

Epperson 09/526,106

=> d que 12

L1 585 SEA FILE=REGISTRY ELDLNSG[KE]ILESFR[SP]EERFPMMS.{0-}PNDERDTT[TM
]PVAMATT/SQSP

L2 0 SEA FILE=REGISTRY L1 AND SQL<=207 ← ^{max} size of fragment

=> d que 13

L3 0 SEA FILE=REGISTRY GEILESFRPEERFP.{0-}DTTMPVAM|GKILESFRSEERFP.{0-}
-}DTTMPVAM|GKILESFRPEERFP.{0-}DTTTPVAM|GEILESFRSEERFP.{0-}DTTMP
VAM|GEILESFRPEERFP.{0-}DTTTPVAM|GKILESFRSEERFP.{0-}DTTTPVAM|GEI
LESFRSEERFP.{0-}DTTTPVAM/SQSP

← mutant positions w/o wild type

=> d que 114

L1 585 SEA FILE=REGISTRY ELDLNSG[KE]ILESFR[SP]EERFPMMS.{0-}PNDERDTT[TM
]PVAMATT/SQSP

L4 78 SEA FILE=HCAPLUS L1

L5 2 SEA FILE=HCAPLUS L4 AND COMPLEMENTATION

L6 156 SEA BALINT R?/AU

L7 137 SEA HER J?/AU

L8 276 SEA L6 OR L7

L9 9 SEA L8 AND COMPLEMENTATION

L10 55396 SEA BETA(A) LACTAMAS?

L11 30 SEA L10(5A) COMPLEMENTATION

L12 23 SEA L11 NOT L9

L13 7 SEA L12 AND FRAGMENT?

L14 9 DUP REM L5 L9 L13 (9 DUPLICATES REMOVED)

=> d ibib abs 114 1-9

L14 ANSWER 1 OF 9 HCAPLUS. COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2003:551735 HCAPLUS

DOCUMENT NUMBER: 139:112705

TITLE: Methods for detection of molecular and protein
interactions by reporter subunit
complementation and its use in functional
genomics and drug screening

INVENTOR(S): Blau, Helen M.; Balint, Robert F.; Wehrman,
Thomas S.; Her, Jeng-horng

PATENT ASSIGNEE(S): The Board of Trustees of the Leland Stanford Junior
University, USA

SOURCE: PCT Int. Appl., 63 pp.

CODEN: PIXXD2

DOCUMENT TYPE: Patent

LANGUAGE: English

FAMILY ACC. NUM. COUNT: 1

PATENT INFORMATION:

| PATENT NO. | KIND | DATE | APPLICATION NO. | DATE |
|---|------|----------|-----------------|----------|
| WO 2003058197 | A2 | 20030717 | WO 2002-US41587 | 20021226 |
| W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM | | | | |

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
PT, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML,
MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.:

US 2001-344757P P 20011226

AB Methods and compns. for detecting mol. interactions, particularly, protein-protein interactions, using at least two inactive, weakly-complementing ss-lactamase fragments are provided. The invention allows detection of such interactions in eukaryotic and mammalian cells, in situ or in vitro. Detection of mol. interactions in mammalian cells is not limited to the nuclear compartment, but can be accomplished in the cytoplasm, cell surface, organelles, or between these entities. Methods provided utilize novel compns. comprising fusion proteins between mols. of interest and inactive, weakly-complementing-ss-lactamase fragments. Assocn. of the mols. of interest brings the corresponding complementary ss-lactamase fragments into close enough proximity for **complementation** to occur and ss-lactamase activity to be obsd. The invention is useful in the study of protein-protein interactions, functional genomics, agonist and antagonist screening and drug discovery.

L14 ANSWER 2 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2003:696387 HCAPLUS
TITLE: Interaction-activated, circularly permuted proteins
INVENTOR(S): Balint, Robert F.; Her, Jeng-Horng
PATENT ASSIGNEE(S): USA
SOURCE: U.S. Pat. Appl. Publ., 49 pp., Cont.-in-part of U.S.
Ser. No. 526,106.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 3
PATENT INFORMATION:

| PATENT NO. | KIND | DATE | APPLICATION NO. | DATE |
|------------------------|------|----------|-------------------|----------|
| US 2003165825 | A1 | 20030904 | US 2001-764163 | 20010116 |
| PRIORITY APPLN. INFO.: | | | US 2000-175968P P | 20000113 |
| | | | US 2000-526106 A2 | 20000315 |

AB Interaction-activated circularly permuted proteins are disclosed that depend for their functional reassembly into the parent protein on the interaction of heterologous polypeptides or other mols. which have been genetically or chem. conjugated to the break-point termini of engineered enzymes. In addn., methods are provided for identifying circularly permuted marker proteins that will optimally reassemble into a functional parent protein, and which are dependent on the assocn. of heterologous interactor domains. The invention is exemplified by circular permutations of a Class A β -lactamase (TEM-1 of E. coli). Circularly permuted marker proteins that comprise mol. interaction-dependent enzymes particularly find use in (1) cell-based sensors for activation or inhibition of metabolic or signal transduction pathways for high-efficiency, (2) high-throughput screening for agonists/antagonists of the target pathway and in high-throughput mapping of pair-wise protein-protein interactions within and between the proteomes of cells, tissues, and pathogenic organisms, and in (3) cell-based screens for high-throughput selection of inhibitors of any protein-protein interaction.

L14 ANSWER 3 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:353582 HCAPLUS

THIS PAGE BLANK (USPTO)

DOCUMENT NUMBER: 136:352286
 TITLE: Affinity maturation by competitive selection
 INVENTOR(S): Balint, Robert F.; Her, Jeng-Horng
 ; Larrick, James W.
 PATENT ASSIGNEE(S): Horizon Biotechnologies, Inc., USA
 SOURCE: PCT Int. Appl., 45 pp.
 CODEN: FIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 1
 PATENT INFORMATION:

| PATENT NO. | KIND | DATE | APPLICATION NO. | DATE |
|---|------|----------|-----------------|----------|
| WO 2002036738 | A2 | 20020510 | WO 2001-US45371 | 20011030 |
| WO 2002036738 | A3 | 20020704 | | |
| W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG AU 2002027112 A5 20020515 AU 2002-27112 20011030 US 2002155502 A1 20021024 US 2001-999413 20011030 EP 1330551 A2 20030730 EP 2001-992763 20011030 R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL, TR | | | | |

PRIORITY APPLN. INFO.: US 2000-245039P P 20001030
 WO 2001-US45371 W 20011030

AB The present invention provides a method of selecting binding pair members with enhanced binding affinity for the cognate binding partner relative to a ref. binding pair member. In particular, the invention provides methods of selecting antibodies with enhanced affinity for an antigen relative to a ref. antibody. This process, "affinity maturation", thereby provides antibodies with superior binding capabilities. Examples show competitive detn. of **complementation** groups among thioredoxin-scaffolded peptides that bind human CD40 and affinity competition between an anti-CD40 antibody and a higher-affinity mutant of the same antibody.

L14 ANSWER 4 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 1

ACCESSION NUMBER: 2002:266601 HCAPLUS
 DOCUMENT NUMBER: 137:29990
 TITLE: Protein-protein interactions monitored in mammalian cells via **complementation** of .beta.-lactamase enzyme fragments
 AUTHOR(S): Wehrman, Tom; Kleaveland, Benjamin; Her, Jeng-Horng; Balint, Robert F.; Blau, Helen M.
 CORPORATE SOURCE: Baxter Laboratory for Genetic Pharmacology, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, 94305, USA
 SOURCE: Proceedings of the National Academy of Sciences of the United States of America (2002), 99(6), 3469-3474
 CODEN: PNASA6; ISSN: 0027-8424
 PUBLISHER: National Academy of Sciences
 DOCUMENT TYPE: Journal

THIS PAGE BLANK (USP)

THIS PAGE BLANK (USP)

LANGUAGE: English

AB We have defined inactive .alpha. and .omega. fragments of .beta.-lactamase that can complement to form a functional enzyme in both bacteria and mammalian cells, serving as a readout for the interaction of proteins fused to the fragments. Crit. to this advance was the identification of a tripeptide, Asn-Gly-Arg, which when juxtaposed at the carboxyl terminus of the .alpha. fragment increased complemented enzyme activity by up to 4 orders of magnitude. .beta.-Lactamase is well suited to monitoring constitutive and inducible protein interactions because it is small (29 kDa), monomeric, and assayable with a fluorescent cell-permeable substrate. The negligible background, the magnitude of induced signal caused by enzymic amplification, and detection of signal within minutes are unparalleled in mammalian protein interaction detection systems published to date.

REFERENCE COUNT: 24 THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 5 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2002:415162 HCAPLUS

DOCUMENT NUMBER: 137:137336

TITLE: A glutamine transport gene, glnQ, is required for fibronectin adherence and virulence of group B streptococci

AUTHOR(S): Tamura, Glen S.; Nittayajarn, Aphakorn; Schoentag, Deborah L.

CORPORATE SOURCE: Children's Hospital and Regional Medical Center and the University of Washington, Seattle, WA, 98105, USA

SOURCE: Infection and Immunity (2002), 70(6), 2877-2885

CODEN: INFIBR; ISSN: 0019-9567

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Group B streptococci (GBS) are a leading cause of neonatal sepsis and meningitis. GBS adhere to fibronectin when it is attached to a solid phase. We isolated a Tn917 transposon mutant, COH1-GT1, which shows decreased adherence to fibronectin. COH1-GT1 also shows decreased adherence to and invasion of respiratory epithelial cells in vitro and decreased virulence in vivo. COH1-GT1 contains a Tn917 insertion in a homolog of glnQ, a gene from Escherichia coli which is required for glutamine transport and codes for a cytoplasmic ATP-binding cassette protein. To confirm that the decreased fibronectin adherence of COH1-GT1 was due to the mutation in glnQ, we constructed COH1-GT2, a strain with a nonpolar site-directed mutation in glnQ. COH1-GT2 showed decreased binding to fibronectin. We also demonstrated that **complementation** of glnQ in trans restored fibronectin adherence to COH1-GT1. COH1-GT1 shows decreased uptake of radiolabeled glutamine and is resistant to the toxic glutamine analog .gamma.-L-glutamylhydrazide, demonstrating that the glnQ gene is required for glutamine transport in GBS. GlnQ lacks a signal sequence and is a cytoplasmic protein in E. coli and thus is unlikely to act as a fibronectin adhesin. GlnQ is transcribed in an operon with a putative glutamine permease gene, glnP, which has a novel predicted structure contg. three distinct domains linked in a single gene. The first two domains are putative glutamine binding domains with homol. to the E. coli periplasmic glutamine binding gene glnH. The third is a putative permease domain with homol. to the E. coli glutamine permease gene glnP. RT-PCR anal. demonstrated that glnP and glnQ are contained within a single transcript. Transcription of scpB, encoding the only known fibronectin-binding adhesin of GBS, is unaffected. We speculate that glnQ may regulate expression of fibronectin adhesins by affecting

THIS PAGE BLANK (USP)

THIS PAGE BLANK (USPTO)

cytoplasmic glutamine levels and that regulation may be posttranscriptional.

REFERENCE COUNT: 33 THERE ARE 33 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 6 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 2

ACCESSION NUMBER: 2002:417208 HCAPLUS

DOCUMENT NUMBER: 137:105986

TITLE: **.beta.-Lactamase** protein
fragment complementation assays as
in vivo and in vitro sensors of protein-protein
interactions

AUTHOR(S): Galarneau, Andre; Primeau, Martin; Trudeau,
Louis-Eric; Michnick, Stephen W.

CORPORATE SOURCE: Departement de Biochimie, Universite de Montreal,
Montreal, QC, H3C 3J7, Can.

SOURCE: Nature Biotechnology (2002), 20(6), 619-622
CODEN: NABIF9; ISSN: 1087-0156

PUBLISHER: Nature America Inc.

DOCUMENT TYPE: Journal

LANGUAGE: English

AB We have previously described a strategy for detecting protein-protein interactions based on protein interaction-assisted folding of rationally designed **fragments** of enzymes. We call this strategy the protein **fragment** complementation assay (PCA). Here we describe PCAs based on the enzyme TEM-1 **.beta.-lactamase** (EC: 3.5.2.6), which include simple colorimetric in vitro assays using the cephalosporin nitrocefin and assays in intact cells using the fluorescent substrate CCF2/AM (ref. 6). Constitutive protein-protein interactions of the GCN4 leucine zippers and of apoptotic proteins Bcl2 and Bad, and the homodimerization of Smad3, were tested in an in vitro assay using cell lysates. With the same in vitro assay, we also demonstrate interactions of protein kinase PKB with substrate Bad. The in vitro assay is facile and amenable to high-throughput modes of screening with signal-to-background ratios in the range of 10:1 to 250:1, which is superior to other PCAs developed to date. Furthermore, we show that the in vitro assay can be used for quant. anal. of a small mol.-induced protein interaction, the rapamycin-induced interaction of FKBP and yeast FRB (the FKBP-rapamycin binding domain of TOR (target of rapamycin)). The assay reproduces the known dissocn. const. and no. of sites for this interaction. The combination of in vitro colorimetric and in vivo fluorescence assays of **.beta.-lactamase** in mammalian cells suggests a wide variety of sensitive and high-throughput large-scale applications, including in vitro protein array anal. of protein-protein or enzyme-protein interactions and in vivo applications such as clonal selection for cells expressing interacting protein partners.

REFERENCE COUNT: 20 THERE ARE 20 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 7 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 2001:904573 HCAPLUS

DOCUMENT NUMBER: 136:50647

TITLE: A protein **fragment** complementation assay
(PCA) for the detection of protein-protein,
protein-small molecule, and protein nucleic acid
interactions based on the E. coli TEM-1
.beta.-lactamase

INVENTOR(S): Michnick, Stephen W.; Galarneau, Andre

PATENT ASSIGNEE(S): Odyssey Pharmaceuticals, Inc., USA

THIS PAGE BLANK (USPTO)

SOURCE: PCT Int. Appl., 29 pp.
 CODEN: PIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 1
 PATENT INFORMATION:

| PATENT NO. | KIND | DATE | APPLICATION NO. | DATE |
|---|------|----------|-----------------|----------|
| WO 2001094617 | A2 | 20011213 | WO 2001-US17886 | 20010601 |
| WO 2001094617 | A3 | 20030206 | | |
| W: AU, CA, JP | | | | |
| RW: AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR | | | | |
| US 2003108869 | A1 | 20030612 | US 2001-870018 | 20010531 |
| EP 1305627 | A2 | 20030502 | EP 2001-983266 | 20010601 |
| R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI, CY, TR | | | | |

PRIORITY APPLN. INFO.:
 US 2000-208485P P 20000602
 US 2001-870018 A 20010531
 WO 2001-US17886 W 20010601

AB The present invention relates generally to protein complementation assays (PCA) and more specifically to PCA assays based on the E. coli TEM-1 .beta.-lactamase for the detection of protein-protein, protein-small mol., and protein nucleic acid interactions. In the present invention, an in vitro colorimetric assay using the substrate, nitrocefin, and an in vivo fluorescence assay using the substrate, CCF2/AM, were disclosed in mammalian cells. The invention is also directed to pos. and neg. survival assays using cephalosporin-cytotoxic prodrug conjugates.

L14 ANSWER 8 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN DUPLICATE 3

ACCESSION NUMBER: 2000:842264 HCAPLUS
 DOCUMENT NUMBER: 134:13994
 TITLE: High-performance enzyme fragment
complementation systems for the identification
 of interacting proteins
 INVENTOR(S): Balint, Robert F.; Her, Jeng-Horng
 PATENT ASSIGNEE(S): Panorama Research, Inc., USA
 SOURCE: PCT Int. Appl., 94 pp.
 CODEN: PIXXD2
 DOCUMENT TYPE: Patent
 LANGUAGE: English
 FAMILY ACC. NUM. COUNT: 3
 PATENT INFORMATION:

| PATENT NO. | KIND | DATE | APPLICATION NO. | DATE |
|---|------|----------|-----------------|----------|
| WO 2000071702 | A1 | 20001130 | WO 2000-US7108 | 20000316 |
| W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM | | | | |
| RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG | | | | |
| EP 1183347 | A1 | 20020306 | EP 2000-946748 | 20000316 |
| R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, | | | | |

THIS PAGE BLANK (USPTO)

IE, SI, LT, LV, FI, RO
 JP 2003500051 T2 20030107 JP 2000-620079 20000316
 PRIORITY APPLN. INFO.: US 1999-135926P P 19990525
 US 2000-175968P P 20000113
 WO 2000-US7108 W 20000316

AB Fragment pairs of a class A .beta.-lactamase (TEM-1 of *Escherichia coli*) are disclosed that depend for their functional reassembly into the parent protein on the interaction of heterologous polypeptides or other mols. which have been genetically or chem. conjugated to the break-point termini of the fragment pairs. In addn., methods are provided for identifying fragment pairs that will optimally reassemble into a functional parent protein. Fragment pairs that comprise mol. interaction-dependent enzymes find use in (1) homogeneous assays and biosensors for any analyte having two or more independent binding sites, (2) tissue-localized activation of therapeutic and imaging reagents in vivo for early detection and treatment of cancer, chronic inflammation, atherosclerosis, amyloidosis, infection, transplant rejection, and other pathologies, (3) cell-based sensors for activation or inhibition of metabolic or signal transduction pathways for high-efficiency, high-throughput screening for agonists/antagonists of the target pathway, (4) high-throughput mapping of pair-wise protein-protein interactions within and between the proteomes of cells, tissues, and pathogenic organisms, (5) rapid selection of antibody fragments or other binding proteins which bind specifically to polypeptides of interest, (6) rapid antigen identification for anti-cell and anti-tissue antibodies, (7) rapid epitope identification for antibodies, (10) cell-based screens for high-throughput selection of inhibitors of any protein-protein interaction.

REFERENCE COUNT: 13 THERE ARE 13 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L14 ANSWER 9 OF 9 HCAPLUS COPYRIGHT 2003 ACS on STN

ACCESSION NUMBER: 1992:525518 HCAPLUS

DOCUMENT NUMBER: 117:125518

TITLE: Nucleotide sequence of the .beta.-lactamase gene from *Enterococcus faecalis* HH22 and its similarity to staphylococcal .beta.-lactamase genes

AUTHOR(S): Zscheck, Karen K.; Murray, Barbara E.

CORPORATE SOURCE: Med. Sch., Univ. Texas, Houston, TX, 77030, USA

SOURCE: Antimicrobial Agents and Chemotherapy (1991), 35(9), 1736-40

CODEN: AMACCQ; ISSN: 0066-4804

DOCUMENT TYPE: Journal

LANGUAGE: English

AB The nucleotide sequence of the constitutively produced .beta.-lactamase (Bla) gene from *E. faecalis* HH22 was shown to be identical to the published sequences of 3 of 4 staphylococcal type A .beta.-lactamase genes; more differences were seen with the genes for staphylococcal type C and D enzymes. One hundred forty nucleotides upstream of the .beta.-lactamase start codon were detd. for an inducible staphylococcal .beta.-lactamase and were identical to those of th constitutively expressed enterococcal gene, indicating that the changes resulting in constitutive expression are not due to changes in the promoter or operator region. Moreover, complementation studies indicated that prodn. of the enterococcal enzyme could be repressed. The genes for the enterococcal Bla and an inducible staphylococcal Bla were each cloned into a shuttle vector and transformed into enterococcal and staphylococcal recipients. The major difference between the backgrounds of the 2 hosts was that more enzyme was produced by the staphylococcal host, regardless of the source of the gene. The location of the enzyme was found to be host dependent,

THIS PAGE BLANK (USPTO)

since cloned gene generated extracellular (free) enzyme in the staphylococcus and cell-bound enzyme in the enterococcus. On the basis of the identities of the enterococcal Bla and several staphylococcal Bla sequences, these data suggest the recent spread of .beta.-lactamase to enterococci and also suggest the loss of a functional repressor.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Db 140 VVGIG 145

RESULT 10
ABG27935
ID ABG27935 standard; Protein; 101 AA.
XX
AC ABG27935;
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27926.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI MPI; 2001-639362/73.
DR N-PSDB; AAS92122.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 58294; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 101 AA;

Query Match 27.2%; Score 366; DB 22; Length 101;
Best Local Similarity 98.6%; Pred. No. 4.9e-30;
Matches 73; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 154 DTTTVAAMATTAKLLTGELLTLASROQLIDMMEADKVAAGPLLRAPAGWFIADKSGAG 213
DB 10 DTTTVAAMATTAKLLTGELLTLASROQLIDMMEADKVAAGPLLRAPAGWFIADKSGAG 69

OY 214 ERGSRGIIAALGPD 227
DB 70 ERGSRGIIAALGPD 83

RESULT 11
ABG27919
ID ABG27919 standard; Protein; 88 AA.
XX
AC ABG27919;
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27910.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
PI MPI; 2001-639362/73.
DR N-PSDB; AAS92106.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 58278; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 88 AA;

Query Match 24.2%; Score 325; DB 22; Length 88;
Best Local Similarity 80.0%; Pred. No. 7.3e-26;
Matches 68; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

OY 4 TLVKKDADQDQAGYIBLDNSGEILSFREERPMWSTFVLGAVLSRIDAQ 63
DB 4 TLVKKDADQDQAGYIBLDNSGEILSFREERPMWSTFVLGAVLSRIDAQ 63

RESULT 14
AAU45138
ID AAV45138 standard; Protein; 158 AA.
XX
AC AAV45138;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #6034.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PE 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAN, Peking DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maineuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
DR N-PSDB: AAS59525.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1: SEQ ID No 6333; 1069pp; English.
XX
CC Sequences AAV39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 158 AA;
Query Match 6.0%; Score 80.5; DB 22; Length 158;
Best Local Similarity 33.3%; Pred. No. 4.5;
Matches 26; Conservative 9; Mismatches 36; Indels 7; Gaps 2;

0Y 110 ANLLTTGSPRELFALFNMGDHWRDRWPELNAIRNDEDDTTPVAMATLR--K 167
DB 80 ARKMTSTSGAPVSSASTARQASPTVGPWEPSTRPAT-----TTSAPLAFRDLRSGR 134
0Y 168 LUTGELLTLASRQQLIDW 185

DB 135 GRMGSVLWLTATRIPLINW 152

RESULT 15
AAW55524
ID AAW55524 standard; Protein; 200 AA.
XX
AC AAW55524;
XX
DT 02-JUL-1998 (first entry)
XX
DE H. pylori ORF 29ep10720_24432762_c3_39 cellular protein.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
OS Helicobacter pylori.
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
PE 27-MAR-1997; 97WO-US05223.
XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTRA) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI: 1997-503122/46.
DR N-PSDB: AAV24933.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claims 14,93; Page 726-727; 1145pp; English.
XX
CC This sequence is a H. pylori cellular protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds.
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 200 AA;
Query Match 5.9%; Score 79; DB 18; Length 200;
Best Local Similarity 25.4%; Pred. No. 9;
Matches 50; Conservative 32; Mismatches 43; Indels 72; Gaps 13;

0Y 22 IELDINSGEILLESF---RSEERPPMSTFRVLLCGAVLSRIDAQEOQLGRRIRHSQNDLV 78


```
Db      11  VYLDTSQKSMETPATIRAEKERPTFSLF-----NRSSGFSDTLKQWVKYENILI 61
Qy      79  ----EYSPVTEKHLDGNTVRELCSAATMSDNTANILL--TTGGPKEL-TAFIHNMG 131
Db      62  DTGGEYSKETQK-----AMLLS-----NIVLVPTT---PSQIDTEVIANML 99
Qy     132  DHYTRLDREPELNEAIPNDERDPTTPVAMATTLKRLT---GELLTLASROQLIDME 187
Db     100  ERLEQLQ---ELNE-----NLRALIVINRMPPTIPTLKEROALIEFIK 138
Qy     188  ----ADKVGPIILRSAL 200
Db     139  ENNPSDKRIT--LLESSL 153
```

Search completed: September 10, 2003, 12:24:53
Job time : 32.7143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_30E_37S_157T
Perfect score: 1345
Sequence: 1 HPETLVKVDADQLGARVQ.....TMDERRQDAIGASLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/pdata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/pdata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/pdata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/pdata/1/1aa/PCUS_COMB.pep:*
6: /cgn2_6/pdata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1 | 74.5 | 5.5 | 124 | US-08-311-731A-202 | Sequence 202, App |
| 2 | 72.5 | 5.4 | 197 | US-09-252-991A-24241 | Sequence 24241, A |
| 3 | 71 | 5.3 | 159 | US-08-991-890-4 | Sequence 4, Appl1 |
| 4 | 69 | 5.1 | 169 | US-09-328-352-4491 | Sequence 4491, Ap |
| 5 | 67.5 | 5.0 | 170 | US-08-858-207A-519 | Sequence 519, App |
| 6 | 66 | 4.9 | 181 | US-08-482-142-195 | Sequence 195, App |
| 7 | 66 | 4.9 | 181 | US-08-478-572-195 | Sequence 195, App |
| 8 | 66 | 4.9 | 181 | US-08-484-296-195 | Sequence 195, App |
| 9 | 65.5 | 4.9 | 158 | US-09-010-809-19 | Sequence 19, Appl |
| 10 | 65 | 4.8 | 150 | US-09-239-909-2 | Sequence 2, Appl1 |
| 11 | 65 | 4.8 | 174 | US-08-557-122A-6 | Sequence 6, Appl1 |
| 12 | 65 | 4.8 | 174 | US-09-262-666-6 | Sequence 6, Appl1 |
| 13 | 65 | 4.8 | 200 | US-08-357-122A-12 | Sequence 12, Appl |
| 14 | 65 | 4.8 | 200 | US-09-262-666-12 | Sequence 12, Appl |
| 15 | 64.5 | 4.8 | 178 | US-09-252-991A-29942 | Sequence 29942, A |
| 16 | 64 | 4.8 | 203 | US-08-624-677A-2 | Sequence 2, Appl1 |
| 17 | 63.5 | 4.7 | 144 | US-08-225-480-4 | Sequence 4, Appl1 |
| 18 | 63.5 | 4.7 | 144 | US-09-118-445-4 | Sequence 4, Appl1 |
| 19 | 62 | 4.6 | 132 | US-09-252-991A-22681 | Sequence 22681, A |
| 20 | 62 | 4.6 | 142 | US-09-252-991A-17611 | Sequence 17611, A |
| 21 | 61.5 | 4.6 | 142 | US-09-345-473E-31 | Sequence 31, Appl |
| 22 | 61 | 4.5 | 189 | US-08-671-548C-48 | Sequence 48, Appl |
| 23 | 60.5 | 4.5 | 103 | US-09-732-210-1282 | Sequence 1282, Ap |
| 24 | 60.5 | 4.5 | 167 | US-08-690-849-2 | Sequence 2, Appl1 |
| 25 | 60.5 | 4.5 | 167 | US-09-004-053-2 | Sequence 2, Appl1 |
| 26 | 60.5 | 4.5 | 198 | US-09-413-814-87 | Sequence 87, Appl |
| 27 | 60 | 4.5 | 144 | US-09-252-991A-25578 | Sequence 25578, A |

| | | | | | |
|----|------|-----|-----|----------------------|-------------------|
| 28 | 60 | 4.5 | 158 | US-09-107-532A-4218 | Sequence 4218, Ap |
| 29 | 59.5 | 4.4 | 146 | US-09-134-001C-5269 | Sequence 5269, Ap |
| 30 | 59.5 | 4.4 | 160 | US-09-252-991A-24737 | Sequence 24737, A |
| 31 | 59 | 4.4 | 171 | US-09-107-532A-3979 | Sequence 3979, Ap |
| 32 | 59 | 4.4 | 180 | US-09-194-905-5 | Sequence 5, Appl1 |
| 33 | 59 | 4.4 | 199 | US-09-252-991A-30363 | Sequence 30363, A |
| 34 | 58.5 | 4.3 | 203 | US-09-252-991A-24921 | Sequence 24921, A |
| 35 | 58 | 4.3 | 72 | US-08-858-207A-412 | Sequence 412, App |
| 36 | 58 | 4.3 | 136 | US-09-252-991A-30622 | Sequence 30622, A |
| 37 | 58 | 4.3 | 168 | US-08-451-947-6 | Sequence 6, Appl1 |
| 38 | 58 | 4.3 | 168 | US-08-424-826A-6 | Sequence 6, Appl1 |
| 39 | 58 | 4.3 | 168 | US-08-928-694-6 | Sequence 6, Appl1 |
| 40 | 58 | 4.3 | 168 | US-08-450-842-6 | Sequence 6, Appl1 |
| 41 | 58 | 4.3 | 168 | US-08-451-390-6 | Sequence 6, Appl1 |
| 42 | 58 | 4.3 | 168 | US-09-081-06850-6 | Sequence 31347, A |
| 43 | 58 | 4.3 | 183 | US-09-252-991A-31347 | Sequence 29325, A |
| 44 | 58 | 4.3 | 187 | US-09-252-991A-29325 | Sequence 17160, A |
| 45 | 58 | 4.3 | 189 | US-09-252-991A-17160 | |

ALIGNMENTS

RESULT 1
US-08-311-731A-202
Sequence 202, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
INVENTOR: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
DIAGNOSTICS AND THERAPEUTICS
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESSES:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE: 530
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
FAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202
Query Match 5.5%; Score 74.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 1.8;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;
QY 130 MCDHV--TRLDREPELN---EALPDERDT---TPVNAATLTKLLTGELTLASNQ 180

Db 12 MGDGIGMEREGRWGTCOPRLRVPGDEPTLGGRASPEDLITL--NLSTPTMHPSPS 69
QY 181 OLIDMMEA-DKAVGPL-----RSALPAGWTADSGAGERSGRTIALGPDGKPSR 232
Db 70 RDDDWVEPFDALOGTAVFATDGDKATMPAVGIGI-----GASTRGS-GILASISPPROPAR 123

RESULT 2

US-09-252-991A-24241
; Sequence 24241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24241
; LENGTH: 197
; TYPE: PRP
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match
Best Local Similarity 23.68; Score 72.5; DB 4; Length 197;
Matches 37; Conservative 23; Mismatches 42; Indels 55; Gaps 7;

QY 94 VRELCSATMSDNTAANLLTTIGGKELTAFLHNMGDHYRLDRMEPELNEAIPNDER 153
Db 12 VALLSARTTSASLASVPLPIATP---TSARPSAGASLT-----PSPIY 55
QY 154 DWTTPVA-MATTLRLKLTGELLTA-----SQOQLDMKADKVGCP----- 194
Db 56 ATTSPLACRACRTRSLCSG--LANANTSTPGSTSR--WSSSNSISAPVSGRSMPIY 110
QY 195 -----LRSALPAGWTADSGAGERSG 216
Db 111 SMAPFARAVSTWSPVITFAIPAMWHSATATATASRG 147

RESULT 3

US-08-991-890-4
; Sequence 4, Application US/08991890
; Patent No. 6114307
; GENERAL INFORMATION:
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Sprugel, Katherine H.
; APPLICANT: Ren, Hong Ping
; APPLICANT: Humes, Jacqueline M.
; APPLICANT: Hoffman, Ross C.
; APPLICANT: Conklin, Darrell C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; NUMBER OF SEQUENCES: 7
; TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sawlsiak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-08-991-890-4

Query Match
Best Local Similarity 23.88; Score 71; DB 3; Length 159;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

QY 49 VLLCGAVLSRIDAGE-----QLGRRIHYSQN-----DYEYSPVTEKH 87
Db 12 LLLCGAVF--VSPSQIHAFQGRHHHHHGGSGAELKCGPFGKHLLSTYCPMPKPT 69
QY 88 LTDGTVRELCSAATMSDNTAANLLTTIGGKELTAFLHNMGDHYRLDRMEPELNEA 147
Db 70 FT-----TTPGWILLES-GRPKEMVSTNNKCGQAL-----CTISEF 105
QY 148 IPNDERDWTTPVAMA-TTLRLKL 169
Db 106 IPNLSPELKRPLSEGOPSLKRTI 128

RESULT 4

US-09-328-352-4491
; Sequence 4491, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQUENCES: 1
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4491
; LENGTH: 169
; TYPE: PRP
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4491

Query Match
Best Local Similarity 22.24; Score 69; DB 4; Length 169;
Matches 36; Conservative 34; Mismatches 60; Indels 32; Gaps 7;

QY 118 GGPKELETAFLHNMGDHYRLDRMEPELNEAIPNDERDT-----TPVAMATTLRLKL 169
Db 5 GGGTITMKLAQON---RDQWVEQIYOYL---DRLTVREIRIVVFTTIFVYVYIVGYS 58
QY 170 TGEELLTLASRQO-----LIDMMEADKVAGPLRSALPAGWTADSGAGER--GSRG 219
Db 59 LMKHSHLAEQOQRLNDLKLDMVYVQSNV-----TKKPANLELDLSGKIORVAQOOG 112
QY 220 ITAALGPDGKPSRIVVITYTGSQATMDERNQIAETGASLTK 261


```

1 APPLICANT: Shaked, Ze'ev
2 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
3 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
4 NUMBER OF SEQUENCES: 207
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
7 STREET: 610 LINCOLN STREET
8 CITY: WALTHAM
9 STATE: MA
10 COUNTRY: USA
11 ZIP: 02154
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: ASCII TEXT
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/484,296
20
21 FILING DATE:
22 CLASSIFICATION: 435
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: 08/445,307
25 FILING DATE: 07 June 1995
26 ATTORNEY/AGENT INFORMATION:
27 NAME: CRAIG, ANNE I.
28 REGISTRATION NUMBER: 32,976
29 REFERENCE/DOCKET NUMBER: 017,605
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (617) 466-6000
32 TELEFAX: (617) 466-6040
33 INFORMATION FOR SEQ ID NO: 195:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 181 amino acids
36 TYPE: amino acid
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39
40 US-08-484-296-195
41
42 Query Match 4.9%; Score 66; DB 3; Length 181;
43 Best Local Similarity 20.5%; Pred. No. 28;
44 Matches 35; Conservative 24; Mismatches 66; Indels 46; Gaps 5
45
46 QY 4 TLVKVDAEQLGRVGT--IELDNGSEILSEFSRSEERPMSTFVLLGAVLSRIDA 61
47 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
48 17 THAKIRVVAATESAYLAVRNTSLDSEQLVDVEYIQHNGVVGSSYRYV----- 65
49
50 QY 62 GQDGLGRIRHYSQ-----NDLVEYSPYTEKHLTDGMVRELCSAATMSDNTAALL 114
51 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
52 DB 66 ARQSCSRPAAOQLEAVEFANQNTKATIEIRASIDGLEV----- 105
53
54 QY 115 TTIGKPELFAFLHNMGDHVTRLDMEDELNAIPND-----ERDTTPVAM 161
55 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
56 DB 106 --ITGKIDDAFRHYDGRGTTIIQRNGYOSINGNAPAEIDLROMRTVTPIRM 154
57
58 RESULT 9
59 US-09-010-809-19
60 Sequence 19, Application US/09010809B
61 Patent No. 6090601
62 GENERAL INFORMATION:
63 APPLICANT: Gustafsson, Claes
64 APPLICANT: Bellach, Mary C.
65 TITLE OF INVENTION: Epochilone Polyketide Synthases and Encoding DNA
66 FILE REFERENCE: 30062-20020.00
67 CURRENT APPLICATION NUMBER: US/09/010,809B
68 CURRENT FILING DATE: 1998-01-22
69 NUMBER OF SEQ ID NOS: 23
70 SOFTWARE: Patentln Ver. 2.0
71 SEQ ID NO 19
72 LENGTH: 158
73 TYPE: PRT
74 ORGANISM: Sorangium cellulosum

```

US-09-010-809-19

Query Match 4.8%; Score 65.5; DB 3; Length 158;
Best Local Similarity 27.6%; Pred. No. 26;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

QY 106 DNFRANLTLT-----TIGPKELTAFLNMGDHT--RDRNRP-----ELNEAIPNDRD 154
DB 21 NHDAKHLITSROGASAPGADYLRSELEALGASVTLAACDVADPRALKDLNDIP----- 75
QY 155 TTPPVAMATTLTKLITGELTLASRQQLIDWMEADKVAAGPLLRSA-----LPAWGF 205
DB 76 SAHPVAAVVHAASVLDGDLGAMSLERT-----DRVFAPKIDAAHHLHQLTQDKPLAAF 129
QY 206 IADKSGAGGSGRG 219
DB 130 ILFSSVAGVLCSSG 143

RESULT 10

US-09-239-909-2
; Sequence 2, Application US/09239909
; Patent No. 6284952
; GENERAL INFORMATION:
; APPLICANT: Kumo Petrochemical Co. Ltd.
; TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
; FILE REFERENCE: P99-2-6
; CURRENT APPLICATION NUMBER: US/09/239,909
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: EP 99300136.1
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: KOPATIN 1.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: G. max calmodulin4 (SCAM4)
US-09-239-909-2

Query Match 4.8%; Score 65; DB 3; Length 150;
Best Local Similarity 27.9%; Pred. No. 27;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

QY 73 SQNDLVESPYTEKHLTDG---MTVREICSAATMSDNTANLILTTI-----GGPRE 122
DB 6 SEEQIVDKREARGLPDKDGCCTVEELAVINSLDNPFEEELQDMISEVDADNGTIE 65
QY 123 LTAFLHNGDHYTRLDRWEPELNEAIPNDRDTPPVAMATTLTKL--TGELLTLASRQ 180
DB 66 FDEFSLMAKKVADTDA--EELKEAFKVFDPKONGYIS--ASELHVMINLGKELNDEEVE 123
QY 181 QUIDMMEAD 189
DB 124 QMI--KEAD 130

RESULT 11

US-08-557-122A-6
; Sequence 6, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mølland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5879664 disk of No. 5879664 disk of No. 5879664th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980, 204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-122A-6

Query Match 4.8%; Score 65; DB 2; Length 174;
Best Local Similarity 23.9%; Pred. No. 34;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 KVDAEDQLGAR-VGYIELDLNGSEILSEFSEERPPMSTRFVLCGAVLSRIDAGQEQ 65
DB 67 KYRQATELKEKNIPLYVDCTEEEBALCRDQVEGYPTIKIFRGL-----DAVKPY 117
QY 66 LGRR-----IHYS-QNDLVESPYTEKHLTDGKTVAELCSAATMSDNTANLILTT 116
DB 118 QGARQTEAIVSYMKOSLPAAVSPVTPENLEIKTKIVIGYIASDDQANDLFTT 174

RESULT 12

US-09-262-666-6
; Sequence 6, Application US/09262666
; Patent No. 6346244
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Mølland
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6346244 disk of No. 6346244 disk of No. 6346244th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,666
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,122
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3980, 204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 174 amino acids
; TYPE: amino acid


```
; LOCATION: (116)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match          4.8%; Score 64.5; DB 4; Length 178;
Best Local Similarity 26.8%; Pred. No. 41;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

OY 110 ANLLLTFTIG---GPKELTAF-----LENGDHYTRLDRWEPELNEAIPNDERDTTPVAMA 162
   | : | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 25 ARVLLVLELAQQRRRPEGLAAHAGLAHHHPGDHVAR-----AGEDLP---RDVDQLGAAA 74
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 163 TTLKKLLTGELL-----TLASROQLIDMMEADKYAGPLLRSAALPAGWFIADKSG 211
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 LRHHLYLAGALOGSEHHEGIGDGTAAAGEAYVG--QDQEVACPYV--GLAGLEPLAVQGD 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 212 AGERGSRGITAALPPDGK 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 A-----LVGVVGQAGK 141
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: September 10, 2003, 12:31:05
Job time : 11.8571 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_30E_37S_157T

Perfect score: 1345

Sequence: 1 HPEPLVKKVKAEDQIGARVQ.....TMDERNQIAIGASLIRKH 263

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------------------------|
| 1 | 457 | 34.0 | 94 | 15 | US-10-102-806-760 Sequence 760, App |
| 2 | 75 | 5.6 | 202 | 15 | US-10-127-816-9 Sequence 9, App1 |
| 3 | 75 | 5.6 | 202 | 15 | US-10-142-717-12 Sequence 12, App1 |
| 4 | 70.5 | 5.2 | 193 | 11 | US-09-805-354-8 Sequence 8, App1 |
| 5 | 70.5 | 5.2 | 193 | 15 | US-10-144-259-8 Sequence 8, App1 |
| 6 | 70 | 5.2 | 206 | 15 | US-10-156-761-13867 Sequence 13867, A |
| 7 | 69.5 | 5.2 | 184 | 15 | US-10-101-464A-765 Sequence 765, App |
| 8 | 69 | 5.1 | 149 | 15 | US-10-156-761-8136 Sequence 8136, App |
| 9 | 67.5 | 5.0 | 149 | 16 | US-10-233-926-4 Sequence 4, App1 |
| 10 | 67.5 | 5.0 | 187 | 10 | US-09-948-018-8 Sequence 8, App1 |
| 11 | 67 | 5.0 | 178 | 10 | US-09-738-626-5680 Sequence 5680, App |
| 12 | 67 | 5.0 | 194 | 15 | US-10-156-761-8623 Sequence 8623, App |
| 13 | 66 | 4.9 | 202 | 15 | US-10-127-816-11 Sequence 11, App1 |
| 14 | 65.5 | 4.9 | 189 | 9 | US-09-815-242-5862 Sequence 5862, App |
| 15 | 65.5 | 4.9 | 189 | 9 | US-09-815-242-12979 Sequence 12979, A |

| | | | | | |
|----|------|-----|-----|----|---------------------------------------|
| 16 | 65.5 | 4.9 | 191 | 9 | US-09-764-870-390 Sequence 390, App |
| 17 | 65.5 | 4.9 | 191 | 15 | US-10-125-540-380 Sequence 390, App |
| 18 | 64.5 | 4.8 | 191 | 15 | US-10-156-761-8978 Sequence 8978, App |
| 19 | 64 | 4.8 | 162 | 10 | US-09-738-626-4796 Sequence 4796, App |
| 20 | 64 | 4.8 | 179 | 10 | US-09-764-868-757 Sequence 757, App |
| 21 | 64 | 4.8 | 179 | 15 | US-10-106-698-4858 Sequence 4858, App |
| 22 | 64 | 4.8 | 202 | 15 | US-10-189-346-12 Sequence 12, App1 |
| 23 | 63.5 | 4.7 | 88 | 15 | US-10-100-252-6 Sequence 6, App1 |
| 24 | 63.5 | 4.7 | 144 | 15 | US-10-131-406-4 Sequence 13460, A |
| 25 | 63.5 | 4.7 | 202 | 15 | US-10-156-761-13460 Sequence 13460, A |
| 26 | 63.5 | 4.7 | 206 | 10 | US-09-738-626-5425 Sequence 5425, App |
| 27 | 63 | 4.7 | 197 | 15 | US-10-300-827-5 Sequence 5, App1 |
| 28 | 63 | 4.7 | 197 | 15 | US-09-882-227-230 Sequence 230, App |
| 29 | 62.5 | 4.6 | 160 | 12 | US-09-882-227-230 Sequence 9507, App |
| 30 | 62.5 | 4.6 | 190 | 5 | US-10-156-761-9507 Sequence 523, App |
| 31 | 62 | 4.6 | 174 | 9 | US-09-864-761-35777 Sequence 523, App |
| 32 | 62 | 4.6 | 185 | 15 | US-10-101-464A-616 Sequence 523, App |
| 33 | 62 | 4.6 | 189 | 9 | US-09-764-870-523 Sequence 523, App |
| 34 | 62 | 4.6 | 189 | 15 | US-10-125-540-523 Sequence 523, App |
| 35 | 62 | 4.6 | 191 | 11 | US-09-927-827-64 Sequence 31, App1 |
| 36 | 61.5 | 4.6 | 196 | 10 | US-09-862-027-31 Sequence 624, App |
| 37 | 61.5 | 4.6 | 196 | 10 | US-09-738-626-6624 Sequence 16, App1 |
| 38 | 60.5 | 4.5 | 202 | 15 | US-10-189-346-16 Sequence 16, App1 |
| 39 | 60 | 4.5 | 190 | 10 | US-09-738-626-4637 Sequence 4637, App |
| 40 | 59.5 | 4.4 | 182 | 15 | US-10-156-761-12056 Sequence 12056, A |
| 41 | 59.5 | 4.4 | 196 | 15 | US-10-219-220-265 Sequence 265, App |
| 42 | 59.5 | 4.4 | 206 | 9 | US-09-741-669-461 Sequence 461, App |
| 43 | 59.5 | 4.4 | 206 | 9 | US-09-815-242-10335 Sequence 10335, A |
| 44 | 59 | 4.4 | 111 | 9 | US-09-867-550-1868 Sequence 1868, App |
| 45 | 59 | 4.4 | 134 | 11 | US-09-768-235B-40 Sequence 40, App1 |

ALIGNMENTS

RESULT 1
US-10-102-806-760
Sequence 760, Application US/10102806
Publication No. US20030054421A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: P4103P1
CURRENT APPLICATION NUMBER: US/10/102,806
PRIOR FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 760
LENGTH: 94
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (80)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760

Query Match 34.0%; Score 457; DB 15; Length 94;
Best Local Similarity 97.8%; Pred. No. 1.5e-39;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 75 NDLYERSPYTEKHLTDGMVRELCSAATMSDNTANLLTTTGGPKETATLHNGDHY 134
DB 1 NDLYERSPYTEKHLTDGMVRELCSAATMSDNTANLLTTTGGPKETATLHNGDHY 60

OY 135 TRLDWEPELNEAIPNDERDITTPVAMATT 164
 DB 61 TRLDWEPELNEAIPNDERDITTPVAMATT 90

RESULT 2

US-10-127-816-9
 ; Sequence 9, Application US/10127816
 ; Publication No. US20030104416A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Klucher, Kevin M.
 ; APPLICANT: Taft, David W.
 ; APPLICANT: Kindsvogel, Wayne R.
 ; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
 ; FILE REFERENCE: 01-17
 ; CURRENT APPLICATION NUMBER: US/10/127, 816
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/285,408
 ; PRIOR FILING DATE: 2001-04-20
 ; PRIOR APPLICATION NUMBER: US 60/286,482
 ; PRIOR FILING DATE: 2001-04-25
 ; PRIOR APPLICATION NUMBER: US 60/341,050
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 60/341,105
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 09/895,834
 ; PRIOR FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/285,424
 ; PRIOR FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-127-816-9

Query Match 5.6%; Score 75; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 13;
 Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGQOLGRIRHYSOND--LVEYSPVTEKHL-----TDGMTVREL---- 97
 DB 17 LLLAAVLTQTQATAPRSPSRRLSPVPAKDCIHQFKSLSPKELQAFKKAKDAIEKRLLEKDL 76
 OY 98 -CSA-----AITS--DNTAANLLTTIGPKRELTAFLHN 129
 DB 77 RCSSHLPFRAMDVKQLQVOERPKALQAEVALTLKVMENMTDSALATILGQPLHTLSHHS 136
 OY 130 MGDHVT-----RLDRWEPELNEAIPNDERDITTPVAMATTIRKLLTGELLTL 176
 DB 137 QLQCTQLQATAPRSPSRRLSRWLHRLQEA-QSKETPGCLEASVTSNLFRLTLRDLKCV 195
 OY 177 ASRQOLI 183
 DB 196 ANGDQCV 202

RESULT 3

US-10-142-717-12
 ; Sequence 12, Application US/10142717
 ; Publication No. US20030104579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Immunex Corporation
 ; APPLICANT: Baum, Peter R.
 ; APPLICANT: Mosley, Bruce A.
 ; APPLICANT: Ketchum, Randal R.
 ; APPLICANT: Taylor, Scott L.
 ; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
 ; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
 ; CURRENT FILING DATE: 2002-05-08
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 12
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-142-717-12

Query Match 5.6%; Score 75; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 13;
 Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGQOLGRIRHYSOND--LVEYSPVTEKHL-----TDGMTVREL---- 97
 DB 17 LLLAAVLTQTQATAPRSPSRRLSPVPAKDCIHQFKSLSPKELQAFKKAKDAIEKRLLEKDL 76
 OY 98 -CSA-----AITS--DNTAANLLTTIGPKRELTAFLHN 129
 DB 77 RCSSHLPFRAMDVKQLQVOERPKALQAEVALTLKVMENMTDSALATILGQPLHTLSHHS 136
 OY 130 MGDHVT-----RLDRWEPELNEAIPNDERDITTPVAMATTIRKLLTGELLTL 176
 DB 137 QLQCTQLQATAPRSPSRRLSRWLHRLQEA-QSKETPGCLEASVTSNLFRLTLRDLKCV 195
 OY 177 ASRQOLI 183
 DB 196 ANGDQCV 202

RESULT 4

US-09-805-354-8
 ; Sequence 8, Application US/09805354
 ; Publication No. US20030078375A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Amnoubt, M. Amin
 ; APPLICANT: Li, Rui
 ; APPLICANT: Xiong, Jian-Ping
 ; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
 ; FILE REFERENCE: 00786-536001
 ; CURRENT APPLICATION NUMBER: US/09/805,354
 ; CURRENT FILING DATE: 2002-06-04
 ; PRIOR APPLICATION NUMBER: US 09/758,493
 ; PRIOR FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: US 60/221,950
 ; PRIOR FILING DATE: 2000-07-31
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 193
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-805-354-8

Query Match 5.2%; Score 70.5; DB 11; Length 193;

Best Local Similarity 24.2%; Pred. No. 36;
 Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;

OY 59 IDAGQOLGRIRHYSONDIVEYSP--VTEKHLTGMVRELCSAATNSDNTAANLLTTT 116
 DB 36 IGPGLQVG-----VYQGEDVYHEFLNDYRSYKDVYEAASHIEQR----- 77
 OY 117 TCGPKELTAFLNMGCHVTRLDKRE-----PELNAIPNDERD 154
 DB 78 -GTEETRTAF--GIEFARSEAFQKGRKAKRMIVITDGEHSDSPLEKVIQOOSERD 132
 OY 155 TTPVAMA 162
 DB 133 NVTTRYAVA 140

RESULT 5


```

; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-948-018-8

Query Match          5.0%; Score 67.5; DB 10; Length 187;
Best Local Similarity 22.4%; Pred. No. 70;
Matches    41; Conservative   22; Mismatches    57; Indels     63; Gaps      9;

QY      112 LLTTTGPRELTAFLHMGDHYTRLDWEPE-----LNDAIPNDERDTTPVANAT- 163
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      11 LLLALLAPPEAS-----QYGRLEWMPDNKCCSSCLQRNGP-----PPCGALETG  58
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      164 -----TLRKLITGELLTASR--QQIDMMEA-----DKVAGP-----L 195
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      59 DTWKESALPLTRRELSLASQPRLDELEVLEELIYLDPPEPGGGMANGTTHLA 118
        : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      196 IASALPAGWTADKSGAGERSRGIIALGPCKRSR--IVYITYGSOATMDERNROIA 253
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      119 AAYGCPAAW-----STPAYSLRPSRSPRLALIENVAAREPASISQLGTHLA 165
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      254 EIG 256
        ::|
DB      166 QLG 168

RESULT 11
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: AMDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PRF
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680

Query Match          5.0%; Score 67; DB 10; Length 178;
Best local similarity 24.8%; Pred. No. 73;
Matches    25; Conservative   17; Mismatches    37; Indels     22; Gaps      4;

QY      171 GELLTASROOLDIM-----MEDKVAGPLLRSALPAGWFIADKSGAGERSGIIIA 223
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```
Db      61 GELFOAAEORGELNFCAGTLEWSTGVDNPL---TLPRHW-----RRNGRLVA 107
Qy      224 LQPDGRP--SRIVVYTTGSOATMDERNROIAEIGASLIXH 262
Db      108 LDQCKRRVARIKALNDALHETHVLIENKILLETTELIAH 148

RESULT 12
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 194;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

Qy      202 AGWFIADKSGAGERSRG---IINALGPDG--KPSRIVVYTTT 239
Db      75 SGFWYTRSGAGERTGAEYRTIAIGPLAVHEPVVAVVVTT 118

RESULT 13
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match
Best Local Similarity 4.9%; Score 66; DB 15; Length 202;
Matches 41; Conservative 23; Mismatches 65; Indels 62; Gaps 8;

Qy      50 LLCGAVLSRIDAGQEQSLGRIRHYSQND--LVESPYTEKHL-----TDGMTVREL----- 97
Db      17 LLAAVILRTQADPPPRATRLPEVAKDCIIAQFKSLSPKELQAFKRAKAIERKLERDM 76
Qy      98 -CSA-----ATMS--DNTAANLLTTTGGPKREYAFIHN 129
Db      77 RCSSHLISRANDKLOLQOYERPKALOAEVALTKVEMENINDSLTTILGQPLHTSHHS 136
Qy      130 MGDHVT-----RDRWEPELNEA---IPNDRDTTPVMAATYLRKLLTGE 172
Db      137 QLQTCLOLQATAPRPKPSRRLSRWLRLQEAQSKETPCLEDSVT-----SNLFOLLARD 191
Qy      173 LLTLASRQOLI 183
Db      192 LKCYASGDQCV 202

RESULT 14
US-09-815-242-5862
; Sequence 5862, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In.
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5862
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5862

Query Match
Best Local Similarity 4.9%; Score 65.5; DB 9; Length 189;
Matches 42; Conservative 32; Mismatches 84; Indels 47; Gaps 8;

Qy      4 TLVKKYDAEDQAGARYGIELDNGSEILSFSEERFPNMTSTFVILGCA----- 54
Db      10 TLIKI--VOEODNQRFKFEITRNIDQVEMLNQO-----IKKICLTGGAGYIAENI 61
```

```
QY      55 -----VLSTRIDAGEOLGRRIHYSONDLVEY-----SPVTEKHLDGTIVELCSAATM 104
        ||| : | : | : ||| : | : | : ||| : | : | : ||| : | : | :
Db      62 NIPAQIFVEFEDPASOGJGLTLKEOCHGDADVIYPANVTGTSLSLHFFDGOSSRRVGIGTGG 121
        ||| : | : | : ||| : | : | : ||| : | : | : ||| : | : | :
QY      105 SDNTANILLETITIGGPKELTAFLHNMGDHVTRRLDRWERPELEINAIPIRNDROTTPPYAMATT 164
        || : | : | : ||| : | : | : ||| : | : | : ||| : | : | :
Db      122 GMIGCGLYLLSQTIDYKQLT----DMAQNGDR-----NTIDLKVRIHKDFEPPI----- 167
        ||| : | : | : ||| : | : | : ||| : | : | : ||| : | : | :
QY      165 LKRLLTGELLTTLASROOLDIMWEAD 189
        || : | : | : ||| : | : | : ||| : | : | : ||| : | : | :
Db      166 ----PGD-LTPANFGVHLHHLDAD 186
```

```

Search completed: September 10, 2003, 12:33:20
Job time : 17.5714 secs

```

```

US-09-815-242-12979
; Sequence 12979, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haasebeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Frawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EXTRA. 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12979
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12979

```

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 4.98; | Score 65.5; | DB 9; | length 189; |
| Best Local Similarity | 20.58; | Pred. No. 1,1e+02; | | |
| Matches 42; | Conservative 32; | Mismatches 84; | Indels 47; | Gaps 8; |

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E_37S_157T
Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARVQ.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query length | DB ID | Description |
|------------|-------|---------|--------------|-------|-------------|
| 1 | 524 | 39.0 | 105 | 2 | JC2566 |
| 2 | 77 | 5.7 | 191 | 2 | S67447 |
| 3 | 76.5 | 5.7 | 200 | 2 | G97064 |
| 4 | 74.5 | 5.5 | 113 | 2 | T45195 |
| 5 | 74.5 | 5.5 | 172 | 2 | A03606 |
| 6 | 74.5 | 5.5 | 184 | 2 | T21126 |
| 7 | 74.5 | 5.5 | 195 | 2 | T36975 |
| 8 | 72 | 5.4 | 149 | 2 | F71252 |
| 9 | 71.5 | 5.3 | 152 | 2 | T36984 |
| 10 | 71 | 5.3 | 153 | 2 | E81708 |
| 11 | 71 | 5.3 | 177 | 2 | A83753 |
| 12 | 71 | 5.3 | 192 | 2 | G83096 |
| 13 | 70 | 5.2 | 108 | 2 | E84217 |
| 14 | 70 | 5.2 | 170 | 2 | AB0192 |
| 15 | 69.5 | 5.2 | 167 | 2 | DB3360 |
| 16 | 68.5 | 5.1 | 145 | 2 | PA4251 |
| 17 | 68.5 | 5.1 | 180 | 2 | C71869 |
| 18 | 68.5 | 5.1 | 190 | 2 | T09136 |
| 19 | 68.5 | 5.1 | 192 | 2 | A83587 |
| 20 | 68 | 5.1 | 177 | 2 | T48420 |
| 21 | 67.5 | 5.0 | 131 | 2 | AD2281 |
| 22 | 66.5 | 4.9 | 116 | 2 | C82906 |
| 23 | 66.5 | 4.9 | 177 | 2 | D90227 |
| 24 | 66.5 | 4.9 | 196 | 2 | H69647 |
| 25 | 66.5 | 4.9 | 198 | 2 | D95285 |
| 26 | 66 | 4.9 | 146 | 2 | C72703 |
| 27 | 65.5 | 4.9 | 42 | 2 | F56978 |
| 28 | 65.5 | 4.9 | 181 | 2 | AB1902 |
| 29 | 65.5 | 4.9 | 193 | 2 | AE0623 |

| | | | | | | |
|----|------|-----|-----|---|--------|---------------------|
| 30 | 65.5 | 4.9 | 197 | 2 | H90211 | conserved hypothet |
| 31 | 65 | 4.8 | 148 | 2 | E75283 | conserved hypothet |
| 32 | 65 | 4.8 | 150 | 2 | T08585 | calmodulin - soybe |
| 33 | 65 | 4.8 | 160 | 1 | E69186 | conserved hypothet |
| 34 | 65 | 4.8 | 168 | 2 | B75498 | conserved hypothet |
| 35 | 65 | 4.8 | 177 | 1 | B43387 | nonstructural prot |
| 36 | 65 | 4.8 | 177 | 2 | JQ1931 | nonstructural prot |
| 37 | 65 | 4.8 | 180 | 2 | H75262 | hypothetical prote |
| 38 | 64.5 | 4.8 | 151 | 2 | D81333 | hypothetical prote |
| 39 | 64.5 | 4.8 | 113 | 2 | D70580 | probable protein-t |
| 40 | 64 | 4.8 | 162 | 2 | AG0769 | probable acetyltra |
| 41 | 64 | 4.8 | 168 | 2 | C85715 | unknown protein en |
| 42 | 64 | 4.8 | 178 | 2 | I40124 | outer surface prot |
| 43 | 64 | 4.8 | 191 | 2 | E95333 | hypothetical prote |
| 44 | 64 | 4.8 | 191 | 2 | AC2133 | iron(III) diclitrat |
| 45 | 64 | 4.8 | 197 | 2 | AC2133 | |

ALIGNMENTS

RESULT 1

JC2566
bia protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 148[128], 81-86, 1994

A>Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A>Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <MBS>

C:Genetics:

A:Gene: bla

C:Superfamily: beta-lactamase I

Query Match 39.0%, Score 524; DB 2; Length 105;

Best Local Similarity 100.0%; Pred. No. 1.1e-36;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 160 | AMATTLRKLTGELTTLASROLLIDMEADKYAGPLIRSRALPAGRTIADKSGGSGSGRC | 219 |
| DB | 2 | AMATTLRKLTGELTTLASROLLIDMEADKYAGPLIRSRALPAGRTIADKSGGSGSGRC | 61 |
| QY | 220 | IIAALGPDGKPSRIYVITTTGSGATMDERNRQIAEIGASLIKHW | 263 |
| DB | 62 | IIAALGPDGKPSRIYVITTTGSGATMDERNRQIAEIGASLIKHW | 105 |

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: T38062; S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996

A:Reference number: 221766

A:Accession: T38062

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:26944; NID:q1217974; PTDN:CA93808.1; PTD:q1217978; GSPDB:G

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.7%; Score 77; DB 2; Length 191;

[illegible]

```

RESULT 3
G97064
spore coat protein CORTC [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R:Nailling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cld
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79306.1; PID:G15024270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1338

```

[illegible]

```

RESULT 4
T45195
hypothetical protein u1756t [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: 216911
A:Accession: T45195
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KE>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

```

| | |
|-------------|-------------------------------------|
| Query Match | 5.58; Score 74.5; DB 2; Length 113; |
|-------------|-------------------------------------|

[illegible]

RESULT 5
AD3606
molybdopterin biosynthesis mog protein [imported] - *Brucella melitensis* (strain 16M)
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
A:Accession: AD3606
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Muijer, C.; Ios, T.; Ivanov
I.; Mazur, M.; Goldstein, E.; Selkoff, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melit*
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3606
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-172 <KUR>
A:Cross-references: GB:AE008918; PIDN:AL54015.1; PID:917984966; GSPDB:GN00191
Experimental source: strain 16M

| | |
|-----------------------|---|
| C: | Superfamily: molybdenum cofactor biosynthesis protein B moab |
| Query Match | 5.58; Score 74.5; DB 2; Length 172; |
| Best Local Similarity | 25.58; Pred. No. 55; |
| Matches | 40; Conservative 14; Mismatches 42; Indels 61; Gaps 9 |
| OY | 83 VTEKHITDGM-TYRELCSAAITMSNDTANILTTIG-GPKELTAFLHNMGDHYRLDRM 140 |
| | : : : : : |
| Db | 39 IYRRTVLPDGCHSVRD-----TLIDLCDTVAODLLITLTGTGS----- 76 |
| OY | 141 EPELNEAIPNDERDRTTPVANATTLIRKLIT--GELLITLASROOLDIMWADKAGAPLLRS 198 |
| | : : : : : : : : : |
| Db | 77 -----PRDE-----TPENKRAVLHKELPFGEQGRVRSLQ----- 107 |
| OY | 199 ALPAGWFIADKSGAGERGSRRIINALGDPDKRSRIIV 235 |
| | : : : : |
| Db | 108 -TPFA--VLISROTAGSRKSFY---LNLPGKPASTIAM 138 |

RESULT 6
T21126
ADP-ribosylation factor homolog F19H8.3 [similarity] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
C:Accession: T21126
R:Steward, C.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19379
A:Accession: T21126
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-184 <MIL>
A:Cross-references: EMBL:Z93378; PIDN:CAE07583.1; GSPDB:GN00020; CESP:F19H8.3
A:Experimental source: clone F19H8
C:Genetics:
A:Gene: CESP:F19H8.3
A:Map position: 2
A:Introns: 143/2
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end, lipoprotein, myristylation, nucleotide binding, p-loop
E:24-31/Region: nucleotide-binding motif A (p-loop)

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: E81708
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of *Chlamydia trachomatis* Moyn and *Chlamydia pneumoniae* AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A:Accession: E81708
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-153 <STO>
 A:Cross-references: GB:AE002305; GB:AE002160; NID:g7190418; PIDN:AAF39236.1; PID:g719041
 A:Experimental source: strain N199 (MOPn)
 C:Genetics:
 A:Gene: TC0378

Query Match 5.3%; Score 71; DB 2; Length 153;
 Best Local Similarity 34.6%; Pred. No. 92;
 Matches 18; Conservative 10; Mismatches 16; Indels 8; Gaps 2;

124 TAFIHMGDHVTFLDMEP--ELNEAIPNDERDTTPVAMATTURKLTGEL 173
 101 TTVIERGEHVTLSEFPPTDNLNMQDOKTPTP-----REMLSGFL 146

RESULT 11

AB3753
 Hypothetical protein BH0825 [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: AB3753
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: AB3650; MUID:20512582; PMID:11058132
 A:Accession: AB3753
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-177 <STO>
 A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04544.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0825
 C:Superfamily: *Bacillus subtilis* hypothetical protein ykka

Query Match 5.3%; Score 71; DB 2; Length 177;
 Best Local Similarity 20.3%; Pred. No. 11e+02;
 Matches 38; Conservative 28; Mismatches 57; Indels 64; Gaps 8;

42 PMWSTFVLLCGAVLSRIDAGEQLGRRIHYSQNDLVESPVTEKHILTDGMTVRELCSAA 101
 11 PEMDTSGLEFYAMBEAND-----RLH-----HLIEDVTEELYYKG 47
 102 ITMSDNTAANLLTTTIGCPKELTAFIHMGDHVTFLD-RMEPEL-NEAIPND----- 151
 48 SIDGENSEMAQL-----NHLTYDVHVRVFRKIGELALPDSLEAEGPM 89
 152 -ERDTTPVAMATTURKLTGELTFLASRQ-----LIDMMDKVAAGPLRSALP 201
 90 VDKGKLPVTVTSLSVOELIEKORYVALTKETQCALHDDDLARIPVEERQATIRKGL- 148

202 AGWFIAD 208
 149 --WIMAD 153

RESULT 12

G83096
 conserved hypothetical protein PA4399 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
 C:Accession: G83096
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
 A:Reference number: AB2950; MUID:20437337; PMID:10984043
 A:Accession: G83096
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-192 <STO>
 A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AA07787.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4399
 C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;
 Best Local Similarity 26.0%; Pred. No. 1.3e+02;
 Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

53 GAVLSRI-----DAGEQL--GRRIHYSQNDLVESPVTEKHILTDGMTVRELCSA----- 100
 2 GNRLSKITRTGDRGETGLAGRRVPRSHPRTEAIGAVDELNSQLGLLELLEARGAMP 61
 101 -----AITMSDNTAANLLTTTIGCPKELTAFIHMGDHVTFLD-----DRMEPEL-----N 145
 62 GLEETVQALAPVQHR-----LPDGLGELAMPETRALDEFVARELSCIDRMNDELGPLKN 116
 146 EAIPINDERDTTPVAMATTURKLTGELTFLASRQQLIDMMDKVAAGPLRSALPAGWF 205
 117 FILPGSR-----PVAQAHCRSLAR-----SAERQCALDQETLEGLRYLNRLSDLIF 168

206 IADNSGAGEGRSGRII--AALGPD 227
 169 VAAARAIARQVAVELMEAAARPD 192

RESULT 13

E84217
 hypothetical protein Yng0594h [imported] - *Halobacterium* sp. NRC-1
 C:Species: *Halobacterium* sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84217
 R:Ng, W.V.; Kennedy, S.P.; Mahapatra, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
 ; Letthauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freltas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Daniels, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
 A:Title: Genome sequence of *Halobacterium* species NRC-1.
 A:Reference number: AB4160; MUID:20504483; PMID:11016950
 A:Accession: E84217
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-108 <STO>
 A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA019105.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: YNG0594H

Query Match 5.2%; Score 70; DB 2; Length 108;
 Best Local Similarity 27.5%; Pred. No. 69;
 Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

44 MSTFVLLCGAVLSRIDAGEQLGRRIHYSQNDLVESPVTEKHILTDGMTVRELCSAAT 103
 1 MDTTVELDQALLARLEATD-----RVFEVFEALLEVDTVLRFRHDDRV-----GSIY 50

104 MSDNTAANLLTTTIGCPKEL-----TAFIHMGDHVTFLD 139
 51 NDDGTDRMARLTVPGDSFLAVEPTSEVAIVDAARTRDR 92

RESULT 14

AB0192
 conserved hypothetical protein YP01575 [imported] - *Yersinia pestis* (strain CO92)

C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
 C:Accession: AB0192
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AB0192
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-170 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90397.1; PID:q15979615; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO1575
 C:Superfamily: Escherichia coli hypothetical protein b0354

Query Match 5.2%; Score 70; DB 2; Length 170;
 Best Local Similarity 21.0%; Pred. No. 1.3e+02;
 Matches 22; Conservative 25; Mismatches 44; Indels 14; Gaps 3;

OY 56 LSRIAGQEOIGRIHYSQNDLVEYSPYTEKHITDGMTRELCSAITMSDNTAANLLIT 115
 DB 70 NKKIDISKGNIG--FNFTDNNLI-----KKIVDKLITQAGLISGRLAIA-----RLVY 115
 OY 116 TICGPELTAFILNMGDHYRLDRMEPLNPAIRNDERDTTPVA 160
 DB 116 DNSGSEFAIITPASVADKIAQRDASTVINSALSGEEDGDPEPYA 160

RESULT 15

D87360
 conserved hypothetical protein CC0895 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: D87360
 R:Merzhan, W.C.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: D87360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-167 <STO>
 A:Cross-references: GB:AE005673; NID:g13422160; PIDN:AAK22880.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0895

Query Match 5.2%; Score 69.5; DB 2; Length 167;
 Best Local Similarity 25.4%; Pred. No. 1.4e+02;
 Matches 36; Conservative 19; Mismatches 56; Indels 31; Gaps 6;

OY 2 PETLVKVD-----EQDGAHVGIYELDN---SGETIESFRSEERFPMSTFK 48
 DB 24 PWTLLIILNDAGVAVRFQWQMDRLGYARNVLAARKLIVSHGVMEYRYSER-PPRHEYW 82
 OY 49 VILCGAVLSRIDAGQEOIGRIHYSQNDLVEYSPYTEKHITDGMTV-----ELCSAIT 103
 DB 83 LTERGLALSPVILITAEWGDNRVYGR---DKSPVLEFRHRTGCAFHPLACEACGQVVD 138
 OY 104 MSDNTAANLLITIGCPKELTA 125
 DB 139 RRDIERA-----GPDLLTA 152

Search completed: September 10, 2003, 12:26:19
 Job time : 11.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 5.57143 Seconds
(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E_37S_157P
Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARVY.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 77 | 5.7 | 191 | YDA4_SCHPO | Q10346 schistosacch |
| 2 | 75 | 5.6 | 201 | RACG_SCHPO | Q9P850 dictyosteli |
| 3 | 72 | 5.4 | 149 | NDK_TREPA | 083974 treponema p |
| 4 | 68.5 | 5.1 | 196 | RAC2_LOTJA | 040220 lotus japon |
| 5 | 68 | 5.1 | 122 | HC3L_THIFE | P80509 thlobacilli |
| 6 | 66.5 | 4.9 | 116 | R8PA_UREPA | Q9PQ10 ureaplasma |
| 7 | 66.5 | 4.9 | 196 | ALKH_BACSU | P50846 b kbg/kdgy |
| 8 | 65.5 | 4.9 | 206 | KTHY_METAC | Q8th89 methanostarc |
| 9 | 65 | 4.8 | 177 | VNSC_RINDK | P35948 rinderpest |
| 10 | 65 | 4.8 | 177 | VNSC_RINDK | Q03349 rinderpest |
| 11 | 63.5 | 4.7 | 144 | YACQ_ECOLI | P27245 escherichia |
| 12 | 63.5 | 4.7 | 185 | YACQ_ECOLI | P38522 escherichia |
| 13 | 63 | 4.7 | 176 | HSLV_THEMA | Q9WV21 thermotoga |
| 14 | 62.5 | 4.6 | 146 | HBG_RABIT | P02099 oryctolagus |
| 15 | 62.5 | 4.6 | 150 | PRUV_SALTY | Q9XDM6 salmonella |
| 16 | 62.5 | 4.6 | 160 | YVDEI_HELPE | Q25700 helicobacte |
| 17 | 62.5 | 4.6 | 172 | YVDEI_HELPE | P31131 escherichia |
| 18 | 62.5 | 4.6 | 182 | PYRE_STRCO | Q9X877 streptomyces |
| 19 | 62.5 | 4.6 | 184 | HRPL_PSEXY | P37929 pseudomonas |
| 20 | 62.5 | 4.6 | 195 | TRPE_THEVO | Q97946 thermoplasma |
| 21 | 62.5 | 4.6 | 200 | RR4_DELNE | Q9F569 pellia nees |
| 22 | 62 | 4.6 | 173 | Y265_BORBU | P17491 rattus norv |
| 23 | 62 | 4.6 | 174 | IRBB_RAT | P17491 rattus norv |
| 24 | 62 | 4.6 | 182 | RR4_BELCH | Q85500 bifidobacte |
| 25 | 62 | 4.6 | 182 | Y861_BIFLO | Q85500 bifidobacte |
| 26 | 62 | 4.6 | 186 | RRP_RICPR | Q9Z608 rickettsia |
| 27 | 62 | 4.6 | 186 | YCEB_SALTY | P40822 salmonella |
| 28 | 61.5 | 4.6 | 152 | YU33_YERPE | Q8ZCF8 yersinia pe |
| 29 | 61.5 | 4.6 | 178 | HSLV_RALSO | Q8ZCF8 yersinia pe |
| 30 | 61 | 4.5 | 121 | SECR_HUMAN | P09683 homo sapien |
| 31 | 61 | 4.5 | 184 | ARL2_DROME | Q06849 drosophila |
| 32 | 61 | 4.5 | 194 | RR4_TRIEN | Q20334 tris ensata |
| 33 | 61 | 4.5 | 200 | TATB_CAOCR | Q9A6T1 caulobacter |

ALIGNMENTS

| RESULT 1 | ID | YDA4_SCHPO | STANDARD | PRT | 191 AA. |
|----------|--|------------|----------|-----|---------|
| AC | Q10346 | | | | |
| DT | 01-OCR-1996 (Rel. 34, Created) | | | | |
| DT | 01-OCR-1996 (Rel. 34, Last sequence update) | | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | | |
| DE | Hypothetical protein C1P12.04c in chromosome I. | | | | |
| GN | SPAC1P12.04C. | | | | |
| OS | Schizosaccharomyces pombe (Fission yeast). | | | | |
| OC | Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; | | | | |
| OC | Schizosaccharomycetales; Schizosaccharomycetaceae; | | | | |
| OC | Schizosaccharomyces. | | | | |
| OX | NCBI_TaxID=4896; | | | | |
| RN | SEQUENCE FROM N. A. | | | | |
| RP | SEQUENCE FROM N. A. | | | | |
| RC | STRAIN=972; | | | | |
| RX | MEDLINE=21848401; PubMed=11859360; | | | | |
| RA | Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., | | | | |
| RA | Sgouras K., Peat N., Hayles J., Baker S., Basham D., Bowman S., | | | | |
| RA | Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., | | | | |
| RA | Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., | | | | |
| RA | Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G., | | | | |
| RA | Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., | | | | |
| RA | James K., Jones L., Jones M., Leather S., McDonald S., McLean J., | | | | |
| RA | Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neil C., | | | | |
| RA | Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E., | | | | |
| RA | Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., | | | | |
| RA | Skelton J., Simmonds M., Squares K., Squares S., Stevens K., | | | | |
| RA | Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., | | | | |
| RA | Woodward J., Voickart G., Aert R., Robben J., Grymoprez B., | | | | |
| RA | Welfens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S., | | | | |
| RA | Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., | | | | |
| RA | Borzum K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M., | | | | |
| RA | Eger P., Zimmermann W., Medler H., Wambolt R., Purrelle B., | | | | |
| RA | Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S., | | | | |
| RA | Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., | | | | |
| RA | Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G., | | | | |
| RA | Daga R.R., Cruzado L., Jimenez J., Sanchez S., del Rey F., Benito J., | | | | |
| RA | Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., | | | | |
| RA | Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., | | | | |
| RA | Shpakowski G.V., Ussery D., Barrell B.G., Nurse P., | | | | |
| RT | The genome sequence of Schizosaccharomyces pombe. | | | | |
| RL | Nature 415:871-880(2002) | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation | | | | |
| CC | at the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| DR | EMBL; 269944; CNA93808.1; - | | | | |
| DR | PIR; T38062; S67447. | | | | |
| DR | GeneDB_Spombe; SPAC1P12.04c; - | | | | |

KW Hypothetical protein
 SQ SEQUENCE 191 AA; 21549 MW; 65555347F0EBED16 CRC64;
 Query Match 5.7%; Score 77; DB 1; Length 191;
 Best Local Similarity 22.6%; Pred. No. 27;
 Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;
 QY 104 MSDNTAAILLTGTPKELTAFLHNGVDHVRIDRNEPELNEAIPNDEROTTTVYAMAT 163
 DB 1 MSYMSHSLMLOQPSGIDKIAIIVN-----VARLD-----PASSKSTAOQVSMIN 46
 QY 164 TLKRLTGLTLTASRQQLIDNMEADKAVGRLRSALPAGWFIADKSGAGRGSGITAA 223
 DB 47 EFRG-----TLRLPGLYKLIVNFRKDSPEPTYSNAINIGYVTE--GLAFLGKQIISI 99
 QY 224 LGPDGKP-----SRIVVYTTGSGQATMDERNRQI----- 252
 DB 100 ----SKPLDEKDLMSRFLWLTTLTITVQLLRKTEDEKREHDLASNLASLPICIHWS 155
 QY 253 AEIGASLIRH 262
 DB 156 VENGAGLHKH 165
 RESULT 2
 RACG_DICDI STANDARD; PRT; 201 AA.
 ID RACG_DICDI STANDARD; PRT; 201 AA.
 AC 09GFSO:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RAS-related protein racg.
 GN RACG.
 OS Dictyostellium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA MEDLINE=21127961; PubMed=11222756;
 RA Rivero F., Diallich H., Glockner G., Noegel A.A.;
 RT "The Dictyostellium discoideum family of Rho-related proteins.";
 RL Nucleic Acids Res. 29:1068-1079(2001).
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF310893; AAC45130.1; -
 DR HSSP: P21181; IAMA.
 DR DictyDb: DD77777; racg.
 DR InterPro: IPR003578; GTPase_Rho.
 DR InterPro: IPR001806; Ras_transfmr.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam: PF00071; ras; 1.
 DR PRINTS: PR00449; RASTRNSFRANG.
 DR SMART: SM00174; RHO; 1.
 DR TIGRFBMS: TIGR00231; small_gtp; 1.
 KW GTP-binding; Prenylation; Lipoprotein.
 FT NE_BIND 10 17 GTP (BY SIMILARITY).
 FT NE_BIND 57 61 GTP (BY SIMILARITY).
 FT NE_BIND 115 118 GTP (BY SIMILARITY).
 FT DOMAIN 32 40 EFFECTOR REGION (POTENTIAL).
 FT LIPID 198 198 GERANYL-GERANYL (BY SIMILARITY).
 SQ SEQUENCE 201 AA; 22813 MW; F6CE586497CC169 CRC64;
 Query Match 5.6%; Score 75; DB 1; Length 201;
 Best Local Similarity 22.6%; Pred. No. 42;

Matches 36; Conservative 20; Mismatches 59; Indels 44; Gaps 7;
 QY 44 MSTFKVLGCVLSTRIDGQGLGRIRHSQDLV-EYSP-VTEKHITDQNTVRELCSAA 101
 DB 1 MRSIKVCVVG-----EGIGKTSMLSTYSNSISNEYOPTVDNST----- 42
 QY 102 ITMSDNTAAILLTGTPKELTAFLHNGVDHVRIDRNEPELNEAIPNDEROTTTVYAMAT 145
 DB 43 LLMINKRPNYSLMDPTAGQEEFSKLRRLSYPTDVFLLCSFLINPSSFSLDSWQVELN 102
 QY 146 EAIPIIDERDTTTPVAMATTLRLTGLL--TLASRQQL 182
 DB 103 ENCPN-----TPIVVGQMDLKSNSVILDRCEKKOL 135
 RESULT 3
 NDK_TREPA STANDARD; PRT; 149 AA.
 ID NDK_TREPA STANDARD; PRT; 149 AA.
 AC 083974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
 DE (Nucleoside-2-P kinase).
 GN NDK OR TP1010.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NICHOUS;
 RA MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst R., Roberts K., Sandusky M., Weldon J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete.";
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
 CC OTHER THAN ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate - ADP +
 CC nucleoside triphosphate.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the NDK family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF001269; AAC65961.1; -
 DR PIR: F71252; F71252.
 DR HSSP: P22887; INPK.
 DR TIGR: TP1010; -
 DR HAMAP: ME_00451; -; 1.
 DR InterPro: IPR001564; NDK.
 DR Pfam: PF00334; NDK; 1.
 DR PRINTS: PR01243; NUCCDPKINASE.
 DR PRODOM: PD001018; NDK; 1.
 DR SMART: SM00562; NDK; 1.
 DR PROSITE: PS00469; NDP_KINASES; 1.
 KW Transferase; Kinase; ATP-binding; Complete proteome.
 FT ACT_SITE 117 117 BY SIMILARITY.
 SQ SEQUENCE 149 AA; 16657 MW; BB329539671EB0BE CRC64;


```

Db      1  MSTARFK-----VTYGDGANGKTMLISTSYNTPEPTDYVP-----TVFDNFS 44
QY      100  AAIPTSDNTAANILLTTTIGSPKE-----LTAR-LHNGDHVTRLDKWEPE 143
Db      45  ANVVV-DCSTVNLGMDTACQEDYNRLPLSTYRGADVFLLFSLISRASYENISKRWIPE 103
QY      144  LNEALPNDERDRTTPVAMATTLKRLLTGELLTL-ASRQOLDMWEADKYA---GPLRSR 159
Db      104  LRHVAP-----TVP-----VAVGTKLIDLREDROYLLIDHPGATPITTAQGEELKRA 149
QY      200  LPAGMFI 206
Db      150  IGAAYVL 156

```

| | | | | |
|----|--|-----------|------|---------|
| | RESULT 5 | | | |
| ID | HC3L_THIPE | STANDARD; | PRT; | 122 AA. |
| AC | P80309; | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | |
| DR | 01-OCT-1996 (Rel. 34, Last sequence update) | | | |
| DV | 01-NOV-1997 (Rel. 35, Last annotation update) | | | |
| DE | Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase) | | | |
| DE | (Fragments). | | | |
| GN | HOXG. | | | |
| OS | Thiobacillus ferrooxidans. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales | | | |
| OC | Acidithiobacillaceae; Acidithiobacillus. | | | |
| OX | NCBI_TaxID=920; | | | |
| RN | [1] | | | |
| RP | SEQUENCE. | | | |
| RC | STRAIN=ATCC 19859; | | | |
| RX | MEDLINE=96241862; PubMed=8661919; | | | |
| RA | Fischer J., Quenmeyer A., Kostka S., Kraft R., Friedrich C.G.; | | | |
| RT | "Purification and characterization of the hydrogenase from | | | |
| RL | Thiobacillus ferrooxidans."; | | | |
| CC | Arch. Microbiol. 165:289-296(1996). | | | |
| CC | -1 CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 -> 2 | | | |
| CC | ferricytochrome c3. | | | |
| CC | -1 CORFACTOR: IRON. | | | |
| CC | -1 CAPTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS | | | |
| CC | ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES. | | | |
| FW | Oxidoreductase: Iron. | | | |
| FT | NON_CONS | 20 | 21 | |
| FT | NON_CONS | 29 | 30 | |
| FT | NON_CONS | 35 | 36 | |
| FT | NON_CONS | 42 | 43 | |
| FT | NON_CONS | 59 | 60 | |
| FT | NON_CONS | 72 | 72 | |

```

Query Match      5.1%; Score 68; DB 1; Length 122;
Best Local Similarity 26.4%; Pred. No. 81;
Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;

QY 19 VGIYELDNLNGSEILESEFSEERFPMMSTFKVLLCGAVLSRIDAGQEOGLGRIRHISQNDLV 78
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 VGRVEGDLDLDSXSLIEFRN-----ATLANFGG-----LG 41

QY 79 EYSPVTEKHLTDGATVRELCSAATMSDNTAANLLTTTGGPRELT 124
      : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 42 KYAPFGTNYEIGVT-----ISGDNDPQAGLVYT-----PREST 75

RESULT 6
RBEA_UREPA
ID RBEA_UREPA STANDARD; PRT; 116 AA.
QC Q9PDH0;

```

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ribosome-binding factor A.
 GN RBFA OR U0321.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
 NCBI_TaxID=134021;
 RX MEDLINE-20500219; PubMed-11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
 RA Caswell G.H., et al. The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum.
 RL Nature 407:757-762(2000).
 CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
 CC with 30S subunits that are part of 70S ribosomes or polysomes).
 CC Essential for efficient processing of 16S rRNA. May interact with
 CC the 5' terminal helix region of 16S rRNA (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
 CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-ch).
 CC -----
 CC EMBL: AE002129; AAF30730.1; -
 DR HAMAP: MF_00003; -; 1.
 DR InterPro: IPR000238; Rib_bind_facts.
 DR Pfam: PF02033; RBFA. 1.
 DR ProDom: PD007327; Rib_bind_facts; 1.
 DR TIGRFAMs: TIGR00082; rbfA; 1.
 DR PROSITE: PS01319; RBFA; FALSE_NEG.
 DR rRNA processing; Complete proteome.
 KW SEQUENCE 116 AA; 13247 MW; A79DCCT1F0547514 CRC64;
 SQ
 Query Match 4.9%; Score 66.5; DB 1; Length 116;
 Best Local Similarity 28.9%; Pred. No. 1e+02;
 Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;
 QY 144 LNEALPDERDTTPVAMATTIRKLITGEL-----LTLASRQQLIDMNE-ADKVAGPIL 196
 DB 18 INNALANEINDKIAKLATYAVR--LSNDLSVAKIFLDAHRKRSMLKLVENVKVSQ-TL 74
 QY 197 RSALPAGW-----FIADKS 210
 DB 75 RSKLAEMTSTYKVPLEKREYIDET 97
 RESULT 7
 ALKH_BACSU STANDARD: PRT; 196 AA.
 ID ALKH_BACSU
 AC P50846;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE KMG/KDPG aldolase (includes: 4-hydroxy-2-oxoglutarate aldolase
 DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase (KMG-aldolase)); 2-
 DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (phospho-2-
 DE dehydro-3-deoxygluconate aldolase) (phospho-2-keto-3-deoxygluconate
 DE aldolase) (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
 DE aldolase)).
 DE KDCG.
 GN Bacillus subtilis.
 OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 CC NCBI_TaxID=1423;
 OX [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-168 / Marburg;
 RX MEDLINE-96349105; PubMed-8760912;
 RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
 RA Serrit P.;
 RT *Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sera and kds loci cloned in a yeast artificial chromosome.*;
 RL Microbiology 142:2005-2016(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-98044033; PubMed-9384377;
 RA Kunst F., Ogatawa N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bessler M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Cepano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Enlian K.D., Errington J., Fabret C., Ferrati E., Galleron N.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grand C.,
 RA Giuseppe C., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Katamata D., Kasahara Y., Kleerr-Blanchard M., Klein C.,
 RA Kodayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Meisel D., Nakai S., Nodack M.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwa A., Oudga B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadleir Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Soldo B.,
 RA Sekiguchi J., Sekowska A., Serrit P., Serrit P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstira P., Tognoni A.,
 RA Tosto V., Uchlyama S., Vandenbol M., Vannier F., Vassartoli A.,
 RA Viati A., Wandut R., Wedler E., Wedler H., Welterner T., Yata K.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT *The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.*;
 RL Nature 390:249-256(1997).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate - pyruvate +
 CC glyoxylate.
 CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate -
 CC pyruvate + D-glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
 CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 CC OF GLYOXYLATE.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE KMG/KDPG ALDOLASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-ch).
 CC -----
 CC EMBL: LA7838; AAB38480.1; -
 CC EMBL: Z99115; CAB14127.1; -
 DR PIR: H69647; H69647.
 DR Subtilist: BG11396; kdgA.
 DR InterPro: IPR000887; Aldase_KDPG_KMG.
 DR Pfam: PF01081; Aldolase; 1.
 DR TIGRFAMs: TIGR01182; ecd; 1.
 DR PROSITE: PS00159; ALDOLASE_KDPG_KMG_1; FALSE_NEG.
 DR PROSITE: PS00160; ALDOLASE_KDPG_KMG_2; 1.
 KW Lyase; Schiff base; Multifunctional enzyme; Complete proteome.
 ACT_SITE 43 43 BY SIMILARITY.
 FT

DB 145 MWILAN 150

RESULT 10

VNSC_RINDR STANDARD; PRT: 177 AA.

ID VNSC_RINDR STANDARD; PRT: 177 AA.

AC 003339;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-FEB-1995 (Rel. 31, Last annotation update)

DE Nonstructural protein C.

GN C.

OS Rinderpest virus (strain RBOK) (RDV).

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.

OX NCBI_TaxID=36409;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE-9315658; PubMed-8429304;

RA Baron M.D., Shella M.S., Barrett T.;

RT "Cloning and sequence analysis of the phosphoprotein gene of

RL Rinderpest virus.";

RL J. Gen. Virol. 74:299-304(1993).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: X68311; CAA8319.1; -

DR EMBL: Z30697; CAA8319.1; -

DR PIR: JQ1931; JQ1931.

DR InterPro: IPR003875; Paramyxovir_NSC.

DR Pfam: PF02725; Paramyxo_NS_C; 1.

KW Nonstructural protein.

SO SEQUENCE 177 AA; 19926 MW; 76D8D46A6D3FB07 CRC64;

Query Match 4.88; Score 65; DB 1; Length 177;

Best Local Similarly 24.68; Pred. No. 2.2e+02;

Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;

QY 93 TVRELCASATMSDNTAANLLTTTCGPKELTAFLLHNGDHTYRLDRWEPELNEAIPNDE 152

DB 49 TIRISASHASQQLDQAKAACLAATVTKLEENTAVKRS-----WEHSL----- 90

QY 153 RDTTTP-----VAMATTLRLKLTGELLTTLASRQQLIDMMEADKVAAPLLRSALPA 202

DB 91 ---VTPOCIAPRSIITMFMITAVKRLRSKMLTSMFNQAL--MWSK-SGEMRNLRFA 144

QY 203 GWETAD 208

DB 145 MWILAN 150

RESULT 11

MARR_ECOLI STANDARD; PRT: 144 AA.

ID MARR_ECOLI STANDARD; PRT: 144 AA.

AC P27245; P76882; P77582;

DT 01-AUG-1992 (Rel. 23, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Multiple antibiotic resistance protein marr.

GN MARR OR SOXO OR CFXB OR INAR OR B1530.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-93186717; PubMed-8383113;

RA Cohen S.P., Haechler H., Levy S.B.;

RT "Genetic and functional analysis of the multiple antibiotic

RL J. Bacteriol. 175:1484-1492(1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE-97426617; PubMed-9278503;

RA Blatter F.R., Plunkett G., Ili, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RX MEDLINE-97251357; PubMed-9097039;

RA Alha H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,

RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,

RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,

RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,

RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takada J.,

RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;

RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome

RL corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

RN [4]

RP CHARACTERIZATION.

RX MEDLINE-94110219; PubMed-8282690;

RA Ariza R.R., Cohen S.P., Bachhawat N., Levy S.B., Dimple B.;

RT "Repressor mutations in the marrAB operon that activate oxidative

RL stress genes and multiple antibiotic resistance in Escherichia

RL coli.";

RL J. Bacteriol. 176:143-148(1994).

RN [5]

RP CHARACTERIZATION.

RX MEDLINE-95286534; PubMed-7768850;

RA Secane A.S., Levy S.B.;

RT "Characterization of MARR, the repressor of the multiple antibiotic

RL resistance (mar) operon in Escherichia coli.";

RL J. Bacteriol. 177:3414-3419(1995).

RN [6]

RP MUTAGENESIS.

RX MEDLINE-20223625; PubMed-10760140;

RA Aleksun M.N., Kim Y.S., Levy S.B.;

RT "Mutational analysis of MARR, the negative regulator of marrAB

RL expression in Escherichia coli, suggests the presence of two regions

RL required for DNA binding.";

RL Mol. Microbiol. 35:1394-1404(2000).

CC -1- FUNCTION: REPRESSOR OF THE MARRAB OPERON WHICH IS INVOLVED IN THE

CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS

CC GENES. BINDS TO THE MARR OPERATOR/PROMOTER SITE.

CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL

CC REGULATORS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: M96235; AAC16394.1; ALT_INIT.

DR EMBL: AE000250; AAC74603.1; ALT_INIT.

DR EMBL: D90795; BAA15212.1; -

DR EMBL: D90796; BAA15220.1; -

DR EMBL: D90797; BAA15232.1; -

DR PIR: B64907; B64907.

DR PDB: 1JGS; 28-DEC-01.

DR Ecogene: EG11435; marr.

```

DR InterPro: IPR000835; HTH_MARR.
DR Pfam: PF01047; MARR; 1.
DR PRINTS: PR00598; HTHMARR.
DR SMART: SM00347; HTH_MARR; 1.
DR PROSITE: PS01117; HTH_MARR; 1.
KW Transcription regulation; DNA-binding; Repressor;
FT Antibiotic resistance; Complete proteome; 3D-structure.
FT MUTAGEN 45 45
FT MUTAGEN 77 77
FT MUTAGEN 123 144
FT MUTAGEN 144 AA; 16065 MW; BE7DF549E24D1D3 CRC64;
SQ SEQUENCE 144 AA; 16065 MW; BE7DF549E24D1D3 CRC64;

Query Match
Best Local Similarity 29.5%; Score 63.5; DB 1; Length 144;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

QY 66 LGRRIHY--SQND--LVEY--SP--VTEKHLTDGWTWVEICSAITMSDNTANLLTTIG 118
DB 14 LGRILHYNOKRDRLNELSLPDITRAOKVLCST--CAACIT----- 56
QY 119 GKELTAFLH-NMGDHTRLDR-----WEDELNEAIPNDERD-----TTTPVAMATTIRK 167
DB 57 -PVELKRVLSVDLGAITRMIDRLVCKGWVERLPN--PNDKRGVLYKLTGGAICGECGQ 113
QY 168 LLTGELTLASROQLDMMEDAKVA--GPIILRSALP 201
DB 114 LVGQDL-----HQELTKNLTADEAVATLEYLTKVLP 144

RESULT 12
YCJC_ECOLI
ID YCJC_ECOLI STANDARD; PRT; 185 AA.
AC P38522; P76839; P77417;
DT 01-OCT-1994 (Rel. 30; Created)
DT 01-NOV-1997 (Rel. 35; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Hypothetical protein ycjC.
GN YCJC OR B1299 OR SF1304.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G., Titt, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=9725137; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakabe S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 65-185 FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=91216440; PubMed=1840553;

```

```

RA Heim R., Strehler E.E.;
RT "Cloning an Escherichia coli gene encoding a protein remarkably
RT similar to mammalian aldehyde dehydrogenases."
RL Gene 99:15-23(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=S. flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [5]
RP IDENTIFICATION.
RC SPECIES=E.coli;
RA Ruid K.E.;
RL Unpublished observations (AUG-1994).
CC -1- SIMILARITY: SOME, TO H. INFLUENZAE HI0659.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000228; AAC74381.1; -.
DR EMBL: D90768; BAA14868.1; -.
DR EMBL: D90767; BAA14859.1; -.
DR EMBL: M38433; -. NOT_ANNOTATED_CDS.
DR EMBL: AE015157; AAN42915.1; -.
DR PIR: F64878; F64878.
DR EcoGene: EG12431; YCJC.
DR InterPro: IPR007113; Cupin_sup.
DR InterPro: IPR001387; HTH_3.
DR Pfam: PF01381; HTH_3; 1.
DR SMART: SM00350; HTH_XRE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 185 AA; 20091 MW; D7D0D3FD794A7768 CRC64;

Query Match
Best Local Similarity 4.7%; Score 63.5; DB 1; Length 185;
Matches 33; Conservative 20; Mismatches 44; Indels 23; Gaps 6;

QY 145 NEAIPNDERPTTPVAMATLRLKLLTGELTLASROQLDMMEDAKVAGPIILRSALPAGW 204
DB 34 HSAISTIEDOKVSA--ISTLQKLKLVGLSLSE-----FSEPEKPPDEQV----- 78
QY 205 FIADKSGAGERSGQITIALGPDGKPSRIWV-----YTTGSQATMDER-NRQIAEIGASL 259
DB 79 -VINODDLIEGSGVSMKLVHNGNPNRTIAMIFETVQGT--TTGERIKHGESEIGTVL 135

RESULT 13
HSLV_THEMA
ID HSLV_THEMA STANDARD; PRT; 176 AA.
AC Q9WY21;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE ATP-dependent protease hslV (EC 3.4.25.-).
GN HSLV OR TM0521.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;

```

```

RX MEDLINE-99287316; PubMed-10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Ulteckack T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Sutton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (by similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE001728; AAD35606.1; -.
CC PIR: G73365; G72365.
CC HSSP: P31059; 1E94.
CC MEROPS: T01.006; -.
CC TIGR: TM0521; -.
CC HAMAP: MF_00248; -.
CC InterPro: IPR001353; Protease_protease.
CC Pfam: PF00227; Proteasome; 1.
CC K01: Hydrolyase; Protease; Complete proteome.
CC ACT_SITE 6 BY SIMILARITY.
CC SEQUENCE 176 AA; 18933 MW; ECE369602A0ADB02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;
Best Local Similarity 28.2%; Pred. No. 3.2e+02;
Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

QY 117 IGGPKELTALHNGDHVTRDRKEPELNEALPNDERTTVPAAATLRLKLTGLTLT 176
DB 42 IGGKVLGAGFAGSVADAMTLTDFEAKLRE-----WGNLTKR 78
QY 177 ASHQQLIDWMEADKVPILRSALPAGFIADK-----SGAGERSRGII-----AAL 224
DB 79 AAVELAKDW-RTDRV-----LR-RLALLLVADKENIFITISNGE-----VTPDDAAAI 127

QY 225 GPDG 228
DB 128 GSGG 131

RESULT 14
HBG_RABIT
ID HBG_RABIT STANDARD; PRT; 146 AA.
AC P02099;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin gamma chain (Beta-3).
GN HBG.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-82053017; PubMed-62771761;
RT The nucleotide sequence of rabbit embryonic globin gene beta 3.*;
RL J. Biol. Chem. 256:11780-11786(1981).
RN (2)

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-89178632; PubMed-2486295;
RA Margot J.B., Demers G.W., Hardison R.C.;
RT Complete nucleotide sequence of the rabbit beta-1-like globin gene
RT cluster. Analysis of intergenic sequences and comparison with the
RT human beta-1-like globin gene cluster.*;
RL J. Mol. Biol. 205:15-40(1989).
CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE
CC GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE
CC HUMAN GAMMA CHAIN GLOBIN, WHICH CODES FOR A FETAL GLOBIN.
CC -1- SUBUNIT: Heterotrimer of two alpha chains and two gamma chains.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- SIMILARITY: Belongs to the globin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M18818; AAA02984.1; -.
CC EMBL: V00883; CAA24252.1; -.
CC PIR: A02417; HBRB3.
CC HSSP: P02100; 1A9W.
CC InterPro: IPR002337; Beta_haem.
CC InterPro: IPR000971; Globin.
CC Pfam: PF00042; globin; 1.
CC PRINTS: PR00814; BETAHAEM.
CC PROSITE: PS01033; GLOBIN; 1.
CC Heme; Oxygen transport; transport; Erythrocyte; Embryo.
CC INIT_MET 0
CC SEQUENCE 146 AA; 16093 MW; EB8D6C1C24DD2D82 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 146;
Best Local Similarity 30.8%; Pred. No. 2.8e+02;
Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6;

QY 35 FSESERFPMSTFKVLICGALVSDAGOLFGRIRHSQNDLYEVSPTKHLTDGMTV 94
DB 3 FTAEEKAITSTWK-----LVYEDAGAAALGL-----LVYV-FWTFQFFSPGL 48
QY 95 RELCSAITMSDNTAANLLTLTTIGPKELTAFLLHMGDHWTRLD 138
DB 49 SS--SSAINGNPKVKAH-----GKKVLTAFA-----GDAYVAVND 79

RESULT 15
PDUV_SALTY
ID PDUV_SALTY STANDARD; PRT; 150 AA.
AC O9XDM6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Propionidol utilization protein pduv.
GN PDUV OR STM2056.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_Taxid=602;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE-99429843; PubMed-10498708;
RA Bobik T.A., Havemann G.D., Busch R.J., Williams D.S., Aldrich H.C.;
RT The propionidol utilization (pdu) operon of Salmonella enterica
RT serovar typhimurium LT2 includes genes necessary for formation of
RT polyphedral organelles involved in coenzyme B(12)-dependent 1, 2-
RT propionidol degradation.*;
RL J. Bacteriol. 181:5967-5975(1999).
RN (2)
RP SEQUENCE FROM N.A.

```

RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
CC -I- INDUCTION: By propanediol.
CC -I- SIMILARITY: BELONGS TO THE EUTP/PDUV FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026270; AAD39020.1; -;
DR EMBL; AE008790; AAL20960.1; -;
DR Styczen; SG10680; pduV.
KM ATP-binding; Complete proteome.
FT NP_BIND 8 ATP (POTENTIAL).
SQ SEQUENCE 150 AA; 16348 MW; 7771229432F97E56 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 150;
Best local Similarity 19.2%; Pred. No. 2.9e+02;
Matches 29; Conservative 20; Mismatches 75; Indels 27; Gaps 4;

QY 52 CGAVLSRIDAGEQLGRIRHYSONDVEYSPYTEKHITDCMTVRELCSAALTMSDNTAAN 111
DB 12 CG-----KTSLSQSLRGBALHKKTAQAEWSPMAID--TPGEYLENRCILYSALTSACEAD 65
QY 112 LLTLTTGGPKELTAFLHNNGDHVTRLDRWEPELNEAIPNDERDPTTPVAMATTIRKILTG 171
DB 66 VIALVLNADAQWSPF-----SPGFTAPNNRPTIGLVTKADLAEPQRISLVA 111
QY 172 ELTLASRQQLIDWMEADRVAGPLRSALPA 202
DB 112 EMLTQAGAQQIF-----ITSALNNSGIDA 135

Search completed: September 10, 2003, 12:20:52
Job time : 6.57143 secs

THIS PAGE BLANK (USPTO)
THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_30E_37S_157T

Perfect score: 1345

Sequence: 1 HPEETLVKVDKEDQIGARVG.....TMDERNRQIAETGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_TREMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 39.0 | 105 | 2 | Q52026 |
| 2 | 524 | 39.0 | 105 | 2 | Q52330 |
| 3 | 499 | 37.1 | 102 | 2 | Q8RTD8 |
| 4 | 488 | 36.3 | 102 | 2 | Q52639 |
| 5 | 464 | 34.5 | 138 | 2 | Q990F3 |
| 6 | 464 | 34.5 | 138 | 2 | Q9AMA1 |
| 7 | 464 | 34.5 | 139 | 2 | Q9AMA2 |
| 8 | 462 | 34.3 | 138 | 2 | Q9AMA0 |
| 9 | 462 | 34.3 | 139 | 2 | Q9AMA9 |
| 10 | 459 | 34.1 | 139 | 2 | Q9AMA3 |
| 11 | 338 | 25.1 | 67 | 2 | Q53553 |
| 12 | 326 | 24.2 | 95 | 2 | Q8DEB5 |
| 13 | 310 | 23.0 | 62 | 2 | Q9UN58 |
| 14 | 301 | 22.4 | 128 | 2 | Q8V0D0 |
| 15 | 251 | 18.7 | 180 | 2 | Q8KVT2 |
| 16 | | | | | Q8KVT2 |

| | | | | | | |
|----|-------|------|-----|----|--------|---------------------|
| 17 | 241 | 17.9 | 48 | 2 | Q9RLH0 | Q9RLH0 proteus mir |
| 18 | 238 | 17.7 | 52 | 2 | Q9R4T2 | Q9R4T2 shigella fl |
| 19 | 197 | 14.6 | 38 | 2 | P97145 | P97145 escherichia |
| 20 | 169.5 | 12.6 | 134 | 2 | Q8VU13 | Q8VU13 staphylococ |
| 21 | 124.5 | 9.3 | 109 | 2 | Q53698 | Q53698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | P97146 | P97146 escherichia |
| 23 | 87 | 6.5 | 100 | 2 | Q93505 | Q93505 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | Q9XB22 | Q9XB22 bacillus ce |
| 25 | 80.5 | 6.0 | 202 | 16 | Q98520 | Q98520 rhizobium 1 |
| 26 | 80 | 5.9 | 198 | 2 | Q9ACM8 | Q9ACM8 streptococ |
| 27 | 76.5 | 5.7 | 200 | 16 | Q97JER | Q97JER clostridiu |
| 28 | 75 | 5.6 | 181 | 16 | Q8D108 | Q8D108 yersinia pe |
| 29 | 74.5 | 5.5 | 113 | 2 | Q49970 | Q49970 mycobacteri |
| 30 | 74.5 | 5.5 | 172 | 16 | Q8YBW3 | Q8YBW3 bruceella me |
| 31 | 74.5 | 5.5 | 172 | 16 | Q8YBW3 | Q8YBW3 bruceella su |
| 32 | 74.5 | 5.5 | 184 | 5 | Q45379 | Q45379 caenorhabdi |
| 33 | 74.5 | 5.5 | 195 | 16 | Q9RIA0 | Q9RIA0 streptomyce |
| 34 | 73 | 5.5 | 153 | 5 | Q8SSZ0 | Q8SSZ0 clona intes |
| 35 | 73 | 5.4 | 131 | 2 | Q9X9H0 | Q9X9H0 yersinia en |
| 36 | 73 | 5.4 | 175 | 17 | Q8YTH8 | Q8YTH8 pyrobaculum |
| 37 | 73 | 5.4 | 201 | 17 | Q8YTH8 | Q8YTH8 pyrobaculum |
| 38 | 72.5 | 5.4 | 145 | 2 | Q05984 | Q05984 methanosa |
| 39 | 72 | 5.4 | 205 | 6 | Q9N275 | Q9N275 ovis aries |
| 40 | 71.5 | 5.3 | 152 | 16 | Q9BJT8 | Q9BJT8 streptomyce |
| 41 | 71.5 | 5.3 | 152 | 16 | Q9RI91 | Q9RI91 streptomyce |
| 42 | 71.5 | 5.3 | 196 | 2 | Q9AH34 | Q9AH34 pseudomonas |
| 43 | 71 | 5.3 | 150 | 10 | Q94IG4 | Q94IG4 nicotiana t |
| 44 | 71 | 5.3 | 153 | 16 | Q9PKT4 | Q9PKT4 chlamydia m |
| 45 | 71 | 5.3 | 177 | 16 | Q9KEM7 | Q9KEM7 bacillus ha |

ALIGNMENTS

RESULT 1

ID Q52026 PRELIMINARY; PRT; 105 AA.
AC Q52026;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DI 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
GN Beta-lactamase (Fragment).
GS B.la.
OS Pseudomonas aeruginosa.
OC Pseudomonas aeruginosa.
OC Plasmid PR01614.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011664; PubMed=7926843;
RA West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;
RT "Construction of Improved Escherichia-Pseudomonas shuttle vectors
RT derived from pUC18/19 and sequence of the region required for their
RT replication in Pseudomonas aeruginosa.";
RL Gene 148:81-86(1994).
DR EMBL: L30112; AAA66058.1; -.
DR HSSP: P00810; 1XP8.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; beta_lactamase_A.
DR Pfam: PF00144; beta_lactamase_1.
DR PRINTS: PR00116; BLACTAMASEA.
KW plasmid.
FT NON_TER.
SQ SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best local similarity 100.0%; Pred. No. 2.0e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMATIRKLITGLTGLTLASRQQLIDNMEADKVDVAGPLLSALPAGFTIDKSGAGRSRG 219
DB 2 AAMATIRKLITGLTGLTLASRQQLIDNMEADKVDVAGPLLSALPAGFTIDKSGAGRSRG 61

OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 263
 |||||
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 105

RESULT 2

ID 052330 PRELIMINARY; PRT; 105 AA.

AC 052330;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Tnl bla protein (Fragment).
 OS Escherichia coli.
 OC Plasmid RK2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]

RP SEQUENCE FROM N.A.
 RK MEDLINE=90264294; PubMed=2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT *The kll-kor regulon of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT korc.*;
 RL J. Bacteriol. 172:3040-3050(1990).
 DR EMBL: M32794; AAA26408.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER

SQ SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;
 FT NON_TER

Query Match 39.0%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 2.6e-37;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTIRKLITGELLTTLASROQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGERSRG 219
 |||||
 DB 2 AMATTIRKLITGELLTTLASROQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 263
 |||||
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 105

RESULT 3

ID 08RTD8 PRELIMINARY; PRT; 145 AA.

AC 08RTD8;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Palasubramaniam S.;
 RT *Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT ceftriaxime-resistant Klebsiella pneumoniae.*;
 RT Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AF467105; AAL75506.1; -.
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER

FT NON_TER .145 145

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194BAC82 CRC64;

Query Match 37.1%; Score 499; DB 2; Length 145;
 Best Local Similarity 68.8%; Pred. No. 5.8e-35;
 Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

OY 100 AATMSDNTANLITTTGGPKELTAFLLHMGDHYTRLDRWEPELDAIPNDERDTTPV 159
 |||||
 DB 7 AAXYSDNSAANLLATVGGPAGLTAFLRQIGDNVTRLDRWETELNEALGCDARDITTPA 66
 OY 160 AMATTIRKLITGELLTTLASROQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGERSRG 219
 |||||
 DB 67 SMAATTIRKLITLTSORLSARSQRLDKMVDVAVGPIIRSVLPAGWFIADKSGAGERSRG 126
 OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 263
 |||||
 DB 127 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 144

RESULT 4

ID 052639 PRELIMINARY; PRT; 102 AA.

AC 052639;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Beta-lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC Plasmid PRO1600.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]

RP SEQUENCE FROM N.A.

RC TRANSPOSON-Tn1;
 RX MEDLINE=94336757; PubMed=8058819;
 RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT *Deletion and transposon mutagenesis and sequence analysis of the pOF
 RT PRO1600 OriR region found in the broad-host-range plasmids of the pOF
 RT series.*;
 RL Plasmid 31:265-274(1994).
 DR EMBL: L22691; AAA98312.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER

SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 3e-34;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTIRKLITGELLTTLASROQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGERSRG 219
 |||||
 DB 2 AMATTIRKLITGELLTTLASROQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 263
 |||||
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNRQIAEIGASLIKHW 144

RESULT 5

ID 0990F3 PRELIMINARY; PRT; 138 AA.

AC 0990F3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-6988, and R/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMMC, Malaysia."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327049; AAK07464.1; -
DR EMBL; AF327051; AAK07466.1; -
DR HSSP; P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1 1
SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 5.3e-32;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTIGPKELTALFHNMGDHTVRLDRWEPELNEAIPNDRDTPPVAMATTLRLKLT 170
DB 10 NLLATVGGPAGTAFRLQIGDVTTRLDRWETELNEALPGDARDTTPASMAATTLRLKLT 69
OY 171 GELLTLASRQQLIDMEADKVAAGPLRSALPAGWFIADKSGAGERSGRIIALGPDGKP 230
DB 70 SQRLSARSQRLQWVDDRVAGPLIRSVLPAGWFIADKTSKRGARIVALLGPNKNA 129
OY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 6

O9AMA1 PRELIMINARY; PRT; 138 AA.
AC O9AMA1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-18/25;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMMC, Malaysia."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327050; AAK07465.1; -
DR HSSP; P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1 1
SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 138;
Best Local Similarity 68.5%; Pred. No. 5.3e-32;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTIGPKELTALFHNMGDHTVRLDRWEPELNEAIPNDRDTPPVAMATTLRLKLT 170
DB 10 NLLATVGGPAGTAFRLQIGDVTTRLDRWETELNEALPGDARDTTPASMAATTLRLKLT 69
OY 171 GELLTLASRQQLIDMEADKVAAGPLRSALPAGWFIADKSGAGERSGRIIALGPDGKP 230
DB 70 SQRLSARSQRLQWVDDRVAGPLIRSVLPAGWFIADKTSKRGARIVALLGPNKNA 129
OY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 7

O9AMA2 PRELIMINARY; PRT; 139 AA.
AC O9AMA2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-918;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMMC, Malaysia."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327048; AAK07463.1; -
DR HSSP; P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1 1
SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

Query Match 34.5%; Score 464; DB 2; Length 139;
Best Local Similarity 68.5%; Pred. No. 5.4e-32;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTIGPKELTALFHNMGDHTVRLDRWEPELNEAIPNDRDTPPVAMATTLRLKLT 170
DB 10 NLLATVGGPAGTAFRLQIGDVTTRLDRWETELNEALPGDARDTTPASMAATTLRLKLT 69
OY 171 GELLTLASRQQLIDMEADKVAAGPLRSALPAGWFIADKSGAGERSGRIIALGPDGKP 230
DB 70 SQRLSARSQRLQWVDDRVAGPLIRSVLPAGWFIADKTSKRGARIVALLGPNKNA 129
OY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 8

O9AMA0 PRELIMINARY; PRT; 138 AA.
AC O9AMA0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-E/98 9-1;
RA Subramaniam G., Navaratnam P.;
RT "SHV gene sequences from ceftriaxide-resistant *Escherichia coli*
RT isolates from UMC, Malaysia."
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF323052; AKK07467.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1 138
FT 138
SQ SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

| | | | | | |
|----|-----------------------|---|--------------------|-------|----------------------------------|
| | Query Match | 34.3% | Score 462; | DB 2; | Length 138; |
| | Best Local Similarity | 67.7%; | Pred. No.-7.9e-32; | | |
| | Matches | 86; | Conservative | 18; | Mismatches 23; Indels 0; Gaps 0; |
| Oy | 111 | NLLLTGIGPKETAFLLHNMGDHYTRLDRWEEBELNEAIPNDERDTTTPVAMATTLRLILT | 170 | | |
| | | : | : | : | : |
| Dd | 10 | NLLLAIVVGAGAGLAPFLRQIGDNTLRDMRETELNEALPGADADTTTPASMAATLRLILT | 69 | | |
| Oy | 171 | GELLTTSROQLIDMENDKYAGPLIRLSALPAGNFIADKSGAGERSSRGIIIMALPGDKP | 230 | | |
| | | : : : | : | : | : |
| Dd | 70 | SQRTLSARSSQQLLOMMYDDRVAGPLRSVLPAWGFTADRTGASKRGARGIIVALLGPNNKA | 129 | | |
| Oy | 231 | SRIVIVY | 237 | | |
| | | : | : | | |
| Dd | 130 | ERIVLVY | 136 | | |

```

RESULT 9
09AM99
ID      09AM99          PRELIMINARY;      PR1;      138 AA.
AC      09AM99;
DT      01-JUN-2001 (TREMBLrel. 17, Created)
DT      01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT      01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      SHV beta-lactamase (Fragment).
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=E/99 4-1;
RA      Subramaniam G., Navaratnam P.;
RT      *SHV gene sequences from ceftazidime-resistant Escherichia coli
RT      isolates from UMCC, Malaysia."
RL      Submitted (DDBJ-2000) to the EMBL/Genbank/DBJ databases.
RL      EMBL: AF327053; AAK07468.1; -.
DR      HSSP: P14557; ISHV.
DR      InterPro: IPR001466; Beta_lactamase.
DR      InterPro: IPR000871; Beta_lactamase_A.
DR      Pfam: PF00144; Beta-lactamase; 1.
DR      PRINTS: PR00118; BLACTAMASEA.
FT      NON_TER      1
FT      NON_TER      138
SQ      SEQUENCE      138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

```

| | | | | | | |
|----|-----------------------|--|--------------------|-------|----------------|------------------|
| | Query Match | 34.3% | Score 462; | DB 2; | Length 138; | |
| | Best Local Similarity | 67.7% | Pred. No. 7.9e-32; | | | |
| | Matches | 86; | Conservative | 18; | Mismatches 23; | Indels 0; Gaps 0 |
| Oy | 111 | NLLTTGGPKETAFIHNAGDHVTRILDRPEELNEAIPNDEBDTTPYAVAMATIRKLIT | 170 | | | |
| | | : : : : : : : : : : | | | | |
| Dd | 10 | NLLATVGSPACLTAFRLQIGDVTTLRDKRETELNEALPPARDDTTPASMAATIRKLIT | 69 | | | |
| Oy | 171 | GELLTIASROCLIDWNEADRVAGPLLSALPACGFIADSCGSGERGRCIIIALGPDPKP | 230 | | | |
| | | : : : : : : : : : : : | | | | |
| Dd | 70 | SQRISARSOROLQNMWDNRVACPLRSVPACGFINDKTASKRGAGIVALLGNPNNA | 129 | | | |

| | | | |
|----|-----|--------|-----|
| Qy | 231 | SRIWVY | 237 |
| | | : | 1 |
| Db | 130 | ERIVLY | 136 |

RESULT 10

```

ID      09AM98:                                PRELIMINARY:                                PRT:      139 AA.
AC      09AM98:
DT      01-JUN-2001 (TEMBLrel. 17, Created)
DT      01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT      01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE      SHV beta-lactamase (Fragment).
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=E/99 3-2;
RA      Subramaniam G., Navaratnam P.;
RT      SHV gene sequences from ceftriaxime-resistant Escherichia coli
RL      isolates from UMKC, Malaysia.";
RL      Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL: AF327054; AAC07469.1; -.
DR      HSSP: P14557; ISHV.
DR      InterPro: IPR001466; Beta_lactamase.
DR      InterPro: IPR000871; Beta_lactamase_A.
DR      Pfam: PF00144; beta-lactamase; 1.
DR      PRINTS: PR00118; BLACTAMASEA.
FT      NON_TER      1
FT      NON_TER      139
SQ      SEQUENCE      139 AA;      15234 MW;      03361A7925442A8 CRC64;

```

[illegible]

```

RESULT 11
Q9AMA3
ID Q9AMA3 PRELIMINARY: PRT: 139 AA.
DC Q9AMA3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=935;
RA Suddaramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftriaxime-resistant Escherichia coli
RT isolates from UMMC, Malaysia.*;
RL submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327047; AAK07462.1; -.
DR HSSP: P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.

```

DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139 139
SQ SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Query Match
Best Local Similarity 67.7%; Pred. No. 1.4e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGKPKETAFIHNKGDHTRLDREPELNEAIPNDERDTTPVAMATTLKRLT 170
DB 10 HLLLATVGGPAGLTAFIROIGDNTTRLDRETELNEALPGARDTTTPASMAATLRLT 69
QY 171 GELLTATASROQLDWMEDAVAGPLRSALPDAGVFIAPKSGSGERSGITIALGPDGXP 230
DB 70 SQRLSRSQRQLQMWVDVAGPLRSVLPAGWFIADKTGASKRGARIVALLGPNMKA 129
QY 231 SRIIVIVY 237
DB 130 ERIIVIVY 136

RESULT 12

Q53553 PRELIMINARY; PRT; 67 AA.

AC Q53553;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OG Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
flexneri strain";
RL Microbios 83:107-117(1995).
DR EMBL: S81098; AAB35839.2; .
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001465; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KM Plasmid.
FT NON_TER 1
FT NON_TER 67
SQ SEQUENCE 67 AA; 7287 MW; 85F8C85B0363F0CB CRC64;

Query Match
Best Local Similarity 98.5%; Pred. No. 1.1e-21;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 55 VLSRIDAGOEQLGRIRHSQNDLVESPTTEKHLTDGMVRELCSAITMSDNTAAILL 114
DB 1 VLSRIDAGOEQLGRIRHSQNDLVESPTTEKHLTDGMVRELCSAITMSDNTAAILL 60
QY 115 TTIGGPK 121
DB 61 TTIGGPK 67

RESULT 13
Q8GDE5 PRELIMINARY; PRT; 95 AA.
AC Q8GDE5;

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.

RA Bello H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Ames S.G.B.;
RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
Chile";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY150585; AAN7730.1; .
FT NON_TER 1
FT NON_TER 95
SQ SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
Best Local Similarity 65.3%; Pred. No. 2e-20;
Matches 62; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 147 AIPNDERDTTPVAMATTLKRLTGLTLLASRQQLDWMEDAVAGPLRSALPDAGWFI 206
DB 1 ALPGDARDTTPASMAATLRLTLTSORLSRSQRQLQMWVDVAGPLRSVLPAGWFI 60
QY 207 ADKSGAGRGSGRGIITIALGPDGKPSRIIVYITGGS 241
DB 61 ADKTGASERGARGIVALLGPNKASRIIVYIYRDS 95

RESULT 14

Q9JUN58 PRELIMINARY; PRT; 62 AA.

AC Q9JUN58;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OG Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
flexneri strain";
RL Microbios 83:107-117(1995).
DR EMBL: S81099; AAB35840.1; .
DR HSSP: P00810; 1XPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KM Plasmid.
FT NON_TER 1
FT NON_TER 62
SQ SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
Best Local Similarity 96.8%; Pred. No. 2.5e-19;
Matches 60; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 132 DHVTRLDREPELNEAIPNDERDTTPVAMATTLKRLTGLTLLASRQQLDWMEDAV 191
DB 1 DHVTRLDREPELNEAIPNDERDTTPVAMATTLKRLTGLTLLASRQQLDWMEDAV 60
QY 192 AG 193

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 32.7143 Seconds

(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_37S

Perfect score: 1345
Sequence: 1 HPETLVKVKADIEDQIGARVG.....TMDERNQIAEIGASLIRKM 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 583 | 43.3 | 196 | 7 AAP60627 | Beta-urogastrone - |
| 2 | 567 | 42.2 | 146 | 22 AAU23221 | Novel human enzyme |
| 3 | 567 | 42.2 | 146 | 22 AAM90870 | Human Immune/haema |
| 4 | 488 | 36.3 | 127 | 18 AAU20440 | H. pylori cytoplas |
| 5 | 473 | 35.2 | 129 | 22 AAU23220 | Novel human enzyme |
| 6 | 473 | 35.2 | 129 | 22 AAM90871 | Human Immune/haema |
| 7 | 473 | 35.2 | 182 | 22 ABG27917 | Novel human diagno |
| 8 | 463 | 34.4 | 94 | 21 AAB59052 | Breast and ovarian |
| 9 | 400 | 29.7 | 159 | 7 AAP60628 | Beta-urogastrone - |

| | | | | | |
|----|------|------|-----|-------------|---------------------|
| 10 | 372 | 27.7 | 101 | 22 ABG27935 | Novel human diagno |
| 11 | 329 | 24.5 | 88 | 22 ABG27919 | Novel human diagno |
| 12 | 195 | 14.5 | 119 | 7 AAP60626 | Beta-urogastrone - |
| 13 | 145 | 10.8 | 51 | 21 AAY92783 | Vtgs-Delta-lactama |
| 14 | 85 | 6.3 | 200 | 18 AAM55524 | H. pylori ORF 29ep |
| 15 | 79.5 | 5.9 | 158 | 22 AAU45138 | Propionibacterium |
| 16 | 78 | 5.8 | 202 | 24 AAE32716 | Human zcyto24 prot |
| 17 | 78 | 5.8 | 202 | 24 AA016283 | Mouse IMX129840-2 |
| 18 | 71 | 5.3 | 159 | 19 AAM69170 | N-terminally tagged |
| 19 | 70 | 5.2 | 15 | 10 AAP98503 | Sequence encoded b |
| 20 | 69.5 | 5.2 | 170 | 22 AAU56330 | Propionibacterium |
| 21 | 69 | 5.1 | 134 | 21 AAG27220 | zeu may's protein f |
| 22 | 69 | 5.1 | 178 | 22 AAG91926 | C glutamicum prote |
| 23 | 69 | 5.1 | 180 | 21 AAG20206 | Arabidopsis thalia |
| 24 | 69 | 5.1 | 184 | 22 AAG81764 | S. epidermidis ope |
| 25 | 67.5 | 5.0 | 194 | 22 AAB25446 | Pinus radiata cell |
| 26 | 67.5 | 5.0 | 187 | 23 AAU97104 | Human MK61 protein |
| 27 | 67 | 5.0 | 101 | 22 AAU66278 | Propionibacterium |
| 28 | 67 | 5.0 | 168 | 22 ABB68717 | Drosophila melanog |
| 29 | 67 | 5.0 | 202 | 22 AAG82686 | S. epidermidis ope |
| 30 | 67 | 5.0 | 202 | 24 AAE32764 | Human ORFX ORF1346 |
| 31 | 66.5 | 4.9 | 177 | 21 AAB41582 | HTRM clone 0258181 |
| 32 | 66.5 | 4.9 | 196 | 21 AAY73344 | Propionibacterium |
| 33 | 66 | 4.9 | 170 | 22 AAU52472 | Human alpha 2 C4 a |
| 34 | 65.5 | 4.9 | 50 | 17 AAM03591 | Partial corn extra |
| 35 | 65.5 | 4.9 | 136 | 21 AAY44229 | Novel human diagno |
| 36 | 65.5 | 4.9 | 152 | 22 ABG19138 | Novel human diagno |
| 37 | 65.5 | 4.9 | 152 | 22 ABG27771 | Novel signal trans |
| 38 | 65 | 4.8 | 179 | 22 AAB7627 | Human colon cancer |
| 39 | 65 | 4.8 | 179 | 22 AAU71192 | S. pneumoniae derl |
| 40 | 65 | 4.8 | 179 | 22 AAG74084 | Novel central nerv |
| 41 | 65 | 4.8 | 183 | 19 AAY85905 | Novel signal trans |
| 42 | 65 | 4.8 | 185 | 22 AAU87344 | Human colon cancer |
| 43 | 65 | 4.8 | 203 | 18 AAM31266 | S. pneumoniae derl |
| 44 | 64.5 | 4.8 | 102 | 23 ABP34479 | Neospora caninum a |
| 45 | 64.5 | 4.8 | 161 | 18 AAY11282 | Human dehydrogenas |

ALIGNMENTS

| | |
|-----------|---|
| RESULT 1 | |
| AAAP60627 | |
| ID | AAAP60627 standard; Protein: 196 AA. |
| XX | |
| AC | AAAP60627; |
| XX | |
| DT | 25-MAR-2003 (updated) |
| XX | |
| DT | 17-JUN-1991 (first entry) |
| XX | |
| DE | Beta-urogastrone - beta-lactamase fusion protein from pUG2101. |
| XX | |
| KW | Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds; |
| XX | |
| XX | fusion protein; beta-lactamase. |
| OS | Synthetic. |
| XX | |
| XX | |
| FT | Key |
| FT | Protein |
| FT | Location/Qualifiers |
| FT | 1..120 |
| FT | /label= beta-lactamase |
| FT | Misc-difference 121..123 |
| FT | /label= adaptor |
| FT | Protein 124..196 |
| FT | /label= Beta-urogastrone |
| XX | |
| PN | DE3523634-A. |
| XX | |
| XX | 09-JAN-1986. |
| PD | |
| XX | |
| PF | 02-JUL-1985; 85DE-3523634. |
| XX | |
| PR | 02-JUL-1984; 84UP-0137691. |
| XX | |

Novel human diagno
Novel human diagno
Beta-urogastrone -
Vtgs-Delta-lactama
H. pylori ORF 29ep
Propionibacterium
Human zcyto24 prot
Mouse IMX129840-2
N-terminally tagged
Sequence encoded b
Propionibacterium
zeu may's protein f
C glutamicum prote
Arabidopsis thalia
S. epidermidis ope
Pinus radiata cell
Human MK61 protein
Propionibacterium
Drosophila melanog
S. epidermidis ope
Human zcyto25 prot
Human ORFX ORF1346
HTRM clone 0258181
Propionibacterium
Human alpha 2 C4 a
Partial corn extra
Novel human diagno
Novel human diagno
Novel central nerv
Novel signal trans
Human colon cancer
S. pneumoniae derl
Novel central nerv
Neospora caninum a
Human dehydrogenas
Streptococcus pneu

PA (EART) EARTH CHEM CO LTD.
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A,
PI Matsushiro S;
XX WPT: 1986-015031/03.
DR N-PSDB: AAN60631.
XX
PI New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PI and transformed cells contg. lt.
XX
PS Disclosure: Page 56-59; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (updated on 25-MAR-2003 to correct PA field.)
CC
XX
SQ Sequence 196 AA:

Query Match 43.3%; Score 583; DB 7; Length 196;
Best Local Similarity 98.3%; Pred. No. 6,8e-52;
Matches 115; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPETLVKVKDAEDQLGARVGIIEIDLSNGKILSEFRSRRPMSTMATFKYLICGAVISRID 60
DB 24 HPETLVKVKDAEDQLGARVGIIEIDLSNGKILSEFRSRRPMSTMATFKYLICGAVISRVD 83
QY 61 AGGEOLGRRIHYSNDLVESVPTKHLTDGKTVELCSAATMSDNTANILITTI 117
DB 84 AGGEOLGRRIHYSNDLVESVPTKHLTDGKTVELCSAATMSDNTANILITTI 140

RESULT 2
AAU23221
ID AAU23221 standard; Protein: 146 AA.
XX
XX AAU23221;
AC XX
XX 18-DEC-2001 (first entry)
DT XX
XX
DE Novel human enzyme polypeptide #307.
XX
XX Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
OS
XX
XX W0200155301-A2.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465566/50.
DR N-PSDB; AAS41091.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
XX diseases
XX
XX Claim 11; SEQ ID NO 1217; 1180bp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the

CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. hemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC A022915-AA023814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIP0
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 146 AA:
SQ
Query Match 42.2%; Score 567; DB 22; Length 146;
Best Local Similarity 97.4%; Pred. No. 2e-50;
Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 HPEFLVKYKADADQAGARYGYEILDNSGKILSFSESEFPMMSTFKVLCGAVLSRID 60
Db 28 HPEFLVKYKADADQAGARYGYEILDNSGKILSFSESEFPMMSTFKVLCGAVLSRID 87
QY 61 AGQEQLGRRIRHYSQNDLYEYSPVTEKHLTDGMYRELCSAITPMSDMTAAILLT 115
Db 88 AGQEQLGRRIRHYSQNDLYEYSPVTEKHLTDGMYRELCXAATMSDNTAAILLT 142
RESULT 3
ID AAM90870 standard; Protein; 146 AA.
XX
AC AAM90870;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen seq ID NO:18463.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN MO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.

SQ Sequence 146 AA;
 Query Match 42.2%; Score 567; DB 22; Length 146;
 Best Local Similarity 97.4%; Pred. No. 2e-50;
 Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 HPELVKVKDAEDQDQARVGYIELDNSGKILSFSEFRPPMSTFKVLLCGAVLSRID 60
 DB 28 HPELVKVKDAEDQDQARVGYIELDNSGKILSFSEFRPPMSTFKVLLCGAVLSRID 87
 OY 61 AGOEOLGRRIHYSONDLVEYSPYTERKHLTDGMTVRELCXAATMSDNTAANLLLT 115
 DB 88 AGOEOLGRRIHYSONDLVEYSPYTERKHLTDGMTVRELCXAATMSDNTAANLLLT 142

RESULT 4
 AAM20440
 ID AAM20440 standard; protein; 127 AA.
 AC AAM20440;
 DT 14-JUL-1997 (first entry)
 XX
 DE H. pylori cytoplasmic protein, 34574062.aa.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome;
 KW replication; transcription; recombination; repair.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 121 /note= "encoded by codon YGC"
 FT Misc-difference 122 /note= "encoded by codon WGG"
 FT
 XX
 PN W09640893-A1.
 PD 19-DEC-1996.
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Meligaerd BL;
 DR WPI: 1997-052306/05.
 DR N-PSDB; AAT67613.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 61: Pages 615-616; 1481pp; English.
 XX
 CC The present sequence is a Helicobacter pylori cytoplasmic protein
 CC involved in genomic replication, transcription, recombination and repair.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds, useful
 CC as potential H. pylori life cycle activators or inhibitors. The genomic
 CC sequence of H. pylori (ATCC 55679) was determined from overlapping
 CC contigs generated by mechanically shearing the bacterial DNA. The
 CC sequences were analysed by computer evaluation. To identify
 CC predicted coding regions defined by computer evaluation. To identify
 CC likely H. pylori antigens for vaccine development, the amino acid
 CC sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be

CC Isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 CC
 SQ Sequence 127 AA;
 Query Match 36.3%; Score 488; DB 18; Length 127;
 Best Local Similarity 99.0%; Pred. No. 2.4e-42;
 Matches 96; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HPELVKVKDAEDQDQARVGYIELDNSGKILSFSEFRPPMSTFKVLLCGAVLSRID 60
 DB 24 HPELVKVKDAEDQDQARVGYIELDNSGKILSFSEFRPPMSTFKVLLCGAVLSRID 83
 OY 61 AGOEOLGRRIHYSONDLVEYSPYTERKHLTDGMTVREL 97
 DB 84 AGOEOLGRRIHYSONDLVEYSPYTERKHLTDGMTVREL 120

RESULT 5
 AAU23220
 ID AAU23220 standard; Protein; 129 AA.
 AC AAU23220;
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human enzyme polypeptide #306.
 XX
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
 KW nephrotropic; anticoagulant.
 XX
 OS Homo sapiens.
 XX
 PN W0200155301-A2.
 PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 26-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.

```
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227209.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241825.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251857.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX MPI; 2001-465566/50.
XX N-PSDB: AAS41090.
XX
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),
XX immunodeficiency disorders (e.g. AIDS) autoimmune disorders
XX (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
XX metabolic disorders (e.g. phenylketonuria), inflammatory disorders
XX (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
XX blood-related disorders (e.g. haemophilia), reproductive disorders
XX (e.g. infertility) and infectious disorders (e.g. Influenza). The
XX polynucleotides of the invention can also be used in gene therapy.
XX AAV2915-AAV23814 represent the novel human enzyme polypeptides of the
XX invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 129 AA;

Query Match 35.2%; Score 473; DB 22; Length 129;
Best Local Similarity 96.9%; Pred. No. 8.9e-41;
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```



```
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251888.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-483426/52.
XX
XX N-PSDB; AAK6352.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Claim 11: SEQ ID NO 18464; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 129 AA:
SQ
Query Match 35.2%; Score 473; DB 22; Length 129;
Best Local Similarity 96.9%; Pred. No. 8.9e-41;
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 HPEITLVKVADEQDQAGRVYIEIDLNSGKILSFRRSEPRFPMMSTFFVLLCGAVLSRID 60
DB 28 HPEITLVKVADEQDQAGRVYIEIDLNSGKILSFRRSEPRFPMMSTFFVLLCGAVLSRID 87
OY 61 AGQEQDQGRRIHYSQNDLVESPYTEKHLTDGMTVRE 96
DB 88 AGQEQDQGRRIHYSQNDLVESPYTEKHLTDGMTXXE 123
RESULT 7
ABG27917
ID ABG27917 standard; Protein: 182 AA.
XX
XX ABG27917;
```

```
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #27908.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAS92104.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID NO 58276; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 182 AA:
SQ
Query Match 35.2%; Score 473; DB 22; Length 182;
Best Local Similarity 87.3%; Pred. No. 1.5e-40;
Matches 96; Conservative 3; Mismatches 7; Indels 4; Gaps 1;
OY 1 HPEITLVKVADEQDQAGRVYIEIDLNSGKILSFRRSEPRFPMMSTFFVLLCGAVLSRID 60
DB 24 HPEITLVKVADEQDQAGRVYIEIDLNSGKILSFRRSEPRFPMMSTFFVLLCGAVLSRID 83
OY 61 AGQEQDQGRRIHYSQNDLVESPYTEKHLTDGMTVRELCSAATYSDNTAA 110
DB 84 AGQEQDQGRRIHYSQNDLVESPYTEKHLTDGMTFR-----LRRSFSDNTAS 129
RESULT 8
AAB59052
ID AAB59052 standard; Protein: 94 AA.
XX
```

XX AAB59052;
 AC
 XX
 DT 27-MAR-2001 (first entry)
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
 XX
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neutrotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 XX MO20005173-A1.
 PN
 XX 21-SEP-2000.
 PD
 XX 08-MAR-2000; 2000WO-US05881.
 PP
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Rosen CA, Ruben SM;
 XX
 XX WPI: 2000-611515/58.
 DR N-PSDB; AAF21955.
 XX
 XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 XX Claim 11; Page 1228; 1299pp; English.
 PS
 XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiac activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis of,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SQ Sequence 94 AA;
 Query Match 34.4%; Score 463; DB 21; Length 94;
 Best Local Similarity 98.9%; Pred. No. 6,1e-40;
 Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 75 NDIVEYSPVTEKHLTDGNTVRELCSAATMSDNTAANLITTTGGKELTAFIHNNGDHV 134
 DB 1 NDIVEYSPVTEKHLTDGNTVRELCSAATMSDNTAANLITTTGGKELTAFIHNNGDHV 60
 OY 135 TRLDWRPELNEALPNDERDRTTTPVAMAT 164
 DB 61 TRLDWRPELNEALPNDERDRTTTPVAMAT 90

RESULT 9
 AAP60628
 ID AAP60628 standard; Protein; 159 AA.
 XX
 AC AAP60628;
 XX
 DT 25-MAR-2003 (updated)
 DT 17-JUN-1991 (first entry)
 XX
 DE Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
 XX
 KW Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
 KW fusion protein; beta-lactamase.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Protein 1..96
 FT /label= beta-lactamase
 FT Misc-difference 97..100
 FT /label= adaptor
 FT Protein 101..159
 FT /label= beta-urogastrone
 XX
 PN DE3523634-A.
 XX
 PD 09-JAN-1986.
 XX
 PE 02-JUL-1985; 85DE-3523634.
 XX
 PR 02-JUL-1984; 84JP-0137691.
 XX
 PA (EART) EARTH CHEM CO LTD.
 XX
 PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
 PI Matsushiro S;
 XX
 DR WPI: 1986-015031/03.
 DR N-PSDB; AAN60632.
 XX
 PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
 PT and transformed cells contg. it.
 XX
 PS Disclosure; Page 59-61; 92pp; German.
 XX
 CC The fusion protein is less easily degraded by proteases and so
 CC protects beta-urogastrone and beta-lactamase collects in the periplasm
 CC of E.coli. It is therefore easy to collect and purify the product.
 CC Beta-urogastrone is the hormone of the salivary glands which suppresses
 CC stomach acid secretion and promotes cell growth, so is useful for
 CC treating ulcers and wounds. Previously the product was obtd. only
 CC in small amts. from human urine.
 CC See also AAN60628, and 30-32.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 159 AA;
 Query Match 29.7%; Score 400; DB 7; Length 159;
 Best Local Similarity 68.3%; Pred. No. 4.3e-33;
 Matches 86; Conservative 8; Mismatches 20; Indels 12; Gaps 2;
 OY 1 HPEITLVKVDADQQLARVGYIELDNSGKILSFSEERPPMNSTFKVLLCGAVLSRID 60
 DB 24 HPEITLVKVDADQQLARVGYIELDNSGKILSFSEERPPMNSTFKVLLCGAVLSRID 83
 OY 61 AGOEDLGRIRIHSNDLVES-----PVTEKHLTDGNTVRELCSAATMSDNTAANL 112
 DB 84 AGOEDLGRIRIHSNDLVESAKRNSDECPLSH---DGYCLHDGVCMTYIALDKYACNC 139
 OY 113 LITTTG 118
 DB 113 LITTTG 118

DB 140 VGVYIG 145

RESULT 10
ABG27935
ID ABG27935 standard; Protein; 101 AA.
XX
AC ABG27935;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27926.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX N-PSDB; AAS92122.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 20; SEQ ID No 58294; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 101 AA;

Query Match 27.7%; Score 372; DB 22; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.7e-30;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 154 DTTMPVAMATTLRLKLTGELTTLASRQQLDWMWADKVPGLRSALPAGFIADKSGAG 213
|||||
DB 10 DTTMPVAMATTLRLKLTGELTTLASRQQLDWMWADKVPGLRSALPAGFIADKSGAG 69

OY 214 ERGSRGITAAALGPD 227
DB 70 ERGSRGITAAALGPD 83

RESULT 11
ABG27919
ID ABG27919 standard; Protein; 88 AA.
XX
AC ABG27919;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27910.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX N-PSDB; AAS92106.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 20; SEQ ID No 58278; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 88 AA;

Query Match 24.5%; Score 329; DB 22; Length 88;
Best Local Similarity 81.2%; Pred. No. 4e-26;
Matches 69; Conservative 2; Mismatches 4; Indels 10; Gaps 1;

OY 4 TLVKVDAEDDGLGARVGYTIELDLSNGKILSFRRSEPRPMSSTFVYLCCGAVLSRIADG 63
|||||

THIS PAGE BLANK (USPTO)

STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202

Query Match 5.4%; Score 72.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 3;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;

QY 130 KGDHV--TLDRPEELN---EATPDERDT---TWVAMATLRKLLTGLTGLTASQ 180
DB 12 MGDGIMERGSRMTGNTQCPLRVYGDSPPLDGRASPEDLTIT--NLSPITMSHPPS 69

QY 181 QLDWMEA-DKVGAPL-----RSALPAGWFIADKSGSGSRGIIALGPDGKPSR 232
DB 70 RDDDWEPFDALDGTAVFDATGDKATPAVGII-----GASTGSG-GILASLSPFRQPAR 123

RESULT 3
US-08-991-890-4
Sequence 4, Application US/08991890
Patent No. 6118307
GENERAL INFORMATION:
APPLICANT: Jaegers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sawlsak, Deborah A.
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-991-890-4

Query Match 5.3%; Score 71; DB 3; Length 159;
Best Local Similarity 23.8%; Pred. No. 6.5;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

QY 49 VLICAVLSRIDAGE-----QLGRIRHSQN-----DVEYSPVTEKH 87
DB 12 LILGAVF--VSPSEIHAEFGRRHHHGGSGAELRGCCPREGKLLSYCPMEKT 69

QY 88 LTDCWTVRELCSAATMSDNTAANLITIGPKELTAFLHNGDHYTRLDRPEELNEA 147
DB 70 FT-----TTGCGWLLS--GRKENVSTSNMKGAL-----GTTSEF 105

QY 148 IPNDERDTMPVAMA--TLRLKL 169
DB 106 IPNLSPELKRPILSEQPSLKITI 128

RESULT 4
US-09-252-991A-24241
Sequence 24241, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24241
LENGTH: 197
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match 5.2%; Score 70.5; DB 4; Length 197;
Best Local Similarity 23.6%; Pred. No. 10;
Matches 37; Conservative 22; Mismatches 43; Indels 55; Gaps 7;

QY 94 VRELCSAATMSDNTAANLITIGPKELTAFLHNGDHYTRLDRPEELNEAIPNDER 153
DB 12 VAKLSARTTSASIASLAVPLPIAPP--TSARFSAGASLT-----PSPIY 55

QY 154 DTTMPVA--MATTLRKLLTGLTGLT-----SRQQLIDMEADKVAGP----- 194
DB 56 ATTSPLACRACRTRSLCSG--LARANSTPGSTRRR--WSSNSISAPVAGSMPIP 110

QY 195 -----LRSALPAGWFIADKSGAGERG 216

```

Db      111  SNAPIARAVSTWSPVITFAPAMHSATATATASSRG 147
          : : : : : : : : : : : : : : : : : :
RESULT 5
US-09-413-814-87
; Sequence 87, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M.
; APPLICANT: Doughtery, Brian A.
; APPLICANT: Goldberg, Steven L.
; APPLICANT: Hoffe, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-87

Query Match
Best Local Similarity 5.2%; Score 70.5; DB 3; Length 198;
Matches 63; Conservative 30; Mismatches 77; Indels 107; Gaps 15;

QY      1  HPELVVKAEDQGLGAVGIELDLSGKILSEFSERFPMSTF-----KYLCCGAV 55
          : : : : : : : : : : : : : : : : : :
Db      7  NPEAVDVVKDKTS--AAAFV---DAFVHKVLEGIGAVYVAIVEGHEFESVLCVAS 60
          : : : : : : : : : : : : : : : : : :
QY      56  LSRIDAGQEQGLGRHYSQNLVEYSPTEKHLTDGMYRELCSAITSDNTANILLT 115
          : : : : : : : : : : : : : : : : : :
Db      61  LVR-----ATRVIOLETBMGRV---GADITMPLAQ-----DILD 114
          : : : : : : : : : : : : : : : : : :
QY      116  TIGGPKELTAFLNHGDVTRLDREPELNEAIPNDERPTMPVAMATTLKILLTGLT 175
          : : : : : : : : : : : : : : : : : :
Db      83  AVG-----ATRVIOLETBMGRV---GADITMPLAQ-----DILD 114
          : : : : : : : : : : : : : : : : : :
QY      176  LASROOLIDMWEADKVAPPLIRSAIPAGMTADK---SGAGER-----GSRGIITAI 225
          : : : : : : : : : : : : : : : : : :
Db      115  LASHYRVVPM---NAHGPLVGQTL-AGSKIRQRIYINIVGVRPHNKRGRDPRLEAPT 169
          : : : : : : : : : : : : : : : : : :
QY      226  PDCKPSRIVYITTSQATM-----DERNQIAELG 256
          : : : : : : : : : : : : : : : : : :
Db      170  PD-----TVIRGDGDTLLVGSDDSDVSREVAEVG 197
          : : : : : : : : : : : : : : : : : :

RESULT 6
US-09-010-809-19
; Sequence 19, Application US/09010809B
; Patent No. 6090601
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; TITLE OF INVENTION: Epithelone polyketide Synthases and Encoding DNA
; FILE REFERENCE: 30062-20020.00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-010-809-19

Query Match
Best Local Similarity 4.9%; Score 65.5; DB 3; Length 158;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

QY      106  DNTANILLT-----TIGGPKELTAFLNHGDVTRLDREPELNEAIPNDERD 154
          : : : : : : : : : : : : : : : : : :
Db      21  NHDKHLITFSROGASAPGADVLRSELEAIGASVFLAACDVADPRALKOLDNIP----- 75
          : : : : : : : : : : : : : : : : : :
QY      155  TTPVAMATITRLKLTGELLTLASROOLIDMWEADKVAPPLIRSA-----LPAGMT 205
          : : : : : : : : : : : : : : : : : :
Db      76  SAHPVAHVHAAVSLDGLGAMSLERI-----DRVFAPKIDAAWHLQUTQDKPLAIF 129
          : : : : : : : : : : : : : : : : : :
QY      206  IADKSGAGERSRG 219
          : : : : : : : : : : : : : : : : : :
Db      130  ILFSSVAGVLCSSG 143
          : : : : : : : : : : : : : : : : : :

RESULT 7
US-09-252-991A-29942
; Sequence 29942, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29942
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (116)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match
Best Local Similarity 4.9%; Score 65.5; DB 4; Length 178;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

QY      110  ANILLTTIG---GPKELTAF-----LHMGSDVTRLDREPELNEAIPNDERPTMPVAMA 162
          : : : : : : : : : : : : : : : : : :
Db      25  AVVLVELLAQORPGLAHAGLAHHRPGDVAR-----AGEDLP---RDYDGLGAAA 74
          : : : : : : : : : : : : : : : : : :
QY      163  TTRKLLTGELL-----TLASROOLIDMWEADKVAPPLIRSAIPAGMTADKSG 211
          : : : : : : : : : : : : : : : : : :
Db      75  LRHHLYLAGALOGEHHEGIGDGTAGAGQAVVC--QDEVAGPRV--GLAGLFLMVGSD 130
          : : : : : : : : : : : : : : : : : :
QY      212  AGERSRGIIALGPDK 229
          : : : : : : : : : : : : : : : : : :
Db      131  A-----LVGVVGQAGK 141
          : : : : : : : : : : : : : : : : : :

RESULT 8
US-09-252-991A-24921
; Sequence 24921, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24921
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24921

Query Match
Best Local Similarity 4.8%; Score 64.5; DB 4; Length 203;
Matches 35; Conservative 16; Mismatches 41; Indels 37; Gaps 7;

OY 135 TRLDWEPEL-----NEAIPNDERDTTPVAMATTLRLKLTGELLTLASHQOLIDME 187
DB 34 TPLDSESESLRVRVSLAMRKAMPMPMAQARMPMKLAEL--RALTGLSTLSSR----DMS- 86
OY 188 ADKVAAPLLRSALPAGNFINDKSGAGERSGRTIAALGPDGKPSRIYIYITGSOATDE 247
DB 87 -----TSAMPFGDI-----SPGWLASTRLEGR-AKLAITATP---AALKV 123
OY 248 RNRQIAEIG 256
DB 124 PTRYSTIG 132

RESULT 9
US-09-239-909-2
; Sequence 2, Application US/09239909
; Patent No. 6284952
; GENERAL INFORMATION:
; APPLICANT: Kuno Petrochemical Co. Ltd.
; TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
; FILE REFERENCE: P99-2-6
; CURRENT APPLICATION NUMBER: US/09/239,909
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: EP 99300136.1
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: KOPATIN 1.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: G. max calmodulin4 (SCAM4)
US-09-239-909-2

Query Match
Best Local Similarity 4.8%; Score 64; DB 3; Length 150;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

OY 73 SONDLYEVSPTTEKHLNDG---MTVRELCSAALITMSDNTAANLLTTI-----GSPKE 122
DB 6 SEEOIVDFKEAFGLFDKDGCTIVEELATVIRSLDONTTEELQDMISEVDADGNGTIE 65
OY 123 LTFVFLHMGDHTVTLDRWEPELNEAIPNDERDTTPVAMATTLRLKLTGELLTLASRQ 180
DB 66 FDEPLSLAKKVKOTDA-EELKEAFKVPDKONGTIS-ASELRHVMINIGKLTDEBEVE 123
OY 181 QLDIMEAD 189
DB 124 QMT--KEAD 130

RESULT 10
US-08-624-677A-2
; Sequence 2, Application US/08624677A
```

```
; Patent No. 6476192
; GENERAL INFORMATION:
; APPLICANT: Lally, Nicola C.
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Dubey, Jitender P.
; TITLE OF INVENTION: Antigens Useful for the serodiagnosis of
; TITLE OF INVENTION: Neosporosis
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Graeter, Janelle S.
; STREET: Rm. 411, Bldg. 005, BARC-W
; CITY: Beltsville
; STATE: MD
; COUNTRY: USA
; ZIP: 20705
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/624,677A
; FILING DATE: 15-APR-1996
; CLASSIFICATION: A35
; ATTORNEY/AGENT INFORMATION:
; NAME: Graeter, Janelle S.
; REGISTRATION NUMBER: 35,024
; REFERENCE/DOCKET NUMBER: 0228.95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-504-6629
; TELEFAX: 301-504-5060
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 203 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-624-677A-2

Query Match
Best Local Similarity 4.8%; Score 64; DB 4; Length 203;
Matches 39; Conservative 23; Mismatches 72; Indels 30; Gaps 7;

OY 54 AYLSRIDAGOEOLGRIRHSONDLYEVSPTTEKHLTGKTVRELCSA--TMSDNTAAN 111
DB 33 AGVSNYDGDGDDAAGNVDD-----SDVTDAITDGEPRVVSQKPRHTOKSLIK 83
OY 112 LITTIIGPKRELFALHNNGDHY-TRLDWEPELNEAIPNDER-DTTPVA----- 160
DB 84 LAVPVVGA--LTSYL--VADRVLPELTSABEGTESIPGKKRKTAVGIALVAALFA 138
OY 161 ---MATTLRLKLTGELLTLASROOLIDMEADKVAAPLLRSALP 201
DB 139 GIGLARTFRHFVPRKSKTYASEDSALGNSSEQYVEGVNGSSDP 182

RESULT 11
US-09-252-991A-24737
; Sequence 24737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24737
```


LENGTH: 160
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24737

Query Match
Best Local Similarity 4.7%; Score 63.5; DB 4; Length 160;
Matches 18; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

3 ETIVKKAEDQDQAGVGYELDLSKIL-----ESPSESRPPMS 45
109 ELLEVKSKFLKSGAALVLEVDENKPAVSFYKSGFEPRFRFVMS 156

RESULT 12

US-08-225-480-4
Sequence 4, Application US/08225480
Patent No. 5817793

GENERAL INFORMATION:

APPLICANT: LEVY, STUART B.

TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/225,480

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/938,085

FILING DATE: 28-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

TELEPHONE/DOCKET NUMBER: T0359/77003

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 144 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-225-480-4

Query Match

Best Local Similarity 4.6%; Score 62.5; DB 2; Length 144;

Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

66 LGRRIHY--SQND--LVEY--SP--VTEKHLTDGMYRELCSAATMSDNTAANLLTTIG 118

14 LGRLLIMVNOKRDLNLEYSPLDITAAQFVLCISIR--CAACIT----- 56

119 GPRELTAFLH--NMGDHYTRLDR-----WEPELNEAIPDERDPTTPVAMATTLKILTG- 171

57 -PVELAKKVLSDVGLALRMDRLVCKGVERLPN--PDKRG-----VLVKLTGTG 104

172 -----ELLTASRQOLLDMWEADRYA--GPLRSALP 201

105 AAICEQHOLVGDHLHQLFKNLTADEVATLEYLLKRYLP 144

RESULT 13
US-09-118-445-4
Sequence 4, Application US/09118445
Patent No. 6391545

GENERAL INFORMATION:

APPLICANT: LEVY, STUART B.

TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESSES:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/118,445

FILING DATE: 17-Jul-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/225,480

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

TELEPHONE/DOCKET NUMBER: T0359/77003

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 144 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-118-445-4

Query Match

Best Local Similarity 4.6%; Score 62.5; DB 4; Length 144;

Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

66 LGRRIHY--SQND--LVEY--SP--VTEKHLTDGMYRELCSAATMSDNTAANLLTTIG 118

14 LGRLLIMVNOKRDLNLEYSPLDITAAQFVLCISIR--CAACIT----- 56

119 GPRELTAFLH--NMGDHYTRLDR-----WEPELNEAIPDERDPTTPVAMATTLKILTG- 171

57 -PVELAKKVLSDVGLALRMDRLVCKGVERLPN--PDKRG-----VLVKLTGTG 104

172 -----ELLTASRQOLLDMWEADRYA--GPLRSALP 201

105 AAICEQHOLVGDHLHQLFKNLTADEVATLEYLLKRYLP 144

US-09-252-991A-22681

Sequence 22681, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 22681
 LENGTH: 132
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-22681

Query Match 4.6%; Score 62; DB 4; Length 132;
 Best Local Similarity 29.8%; Pred. No. 48;
 Matches 28; Conservative 8; Mismatches 34; Indels 24; Gaps 4;

OY 144 LNEATPNDERTTTPYAMATTLKRLTGLTLLASRQOLIMMEADKYAGPLRSALP-- 201
 DB 34 LKQGIIPN--HAGRPQOATVLPMEPNGE-----RRVAFPMKSPPPPR 74
 OY 202 AGMTADKSGAGERSGIIA---ALGPDGKPSR 232
 DB 75 APEVYADRAGPPGPTGCCARRRRAAPAGVPAR 108

RESULT 15
 US-09-345-473E-31
 Sequence 31, Application US/09345473E
 Patent No. 6558903
 GENERAL INFORMATION:
 APPLICANT: Hodge, Martin
 TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof
 FILE REFERENCE: 35800/183781
 CURRENT APPLICATION NUMBER: US/09/345,473E
 CURRENT FILING DATE: 1999-06-30
 NUMBER OF SEQ ID NOS: 62
 SOFTWARE: FASTSEQ for Windows Version 4.0
 SEQ ID NO 31
 LENGTH: 142
 TYPE: PRT
 ORGANISM: Rattus norvegicus
 US-09-345-473E-31

Query Match 4.6%; Score 61.5; DB 4; Length 142;
 Best Local Similarity 41.7%; Pred. No. 61;
 Matches 15; Conservative 5; Mismatches 11; Indels 5; Gaps 1;

OY 210 SGAGERSGIIAALGP-----DGKPSRIYIYTTG 240
 DB 47 SGSGSRDSGGLAVLAPLNDGKASRSDVYSFG 82

Search completed: September 10, 2003, 12:31:03
 Job time: 12.8571 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_37S

Perfect score: 1345

Sequence: 1 HPEPLVKKVDAEDQLGARVY.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*

3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep:*

13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubppa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 463 | 34.4 | 94 | US-10-102-806-760 | Sequence 760, App |
| 2 | 78 | 5.8 | 202 | US-10-127-816-9 | Sequence 9, Appl |
| 3 | 78 | 5.8 | 202 | US-10-142-717-12 | Sequence 12, Appl |
| 4 | 69.5 | 5.2 | 149 | US-10-233-926-4 | Sequence 4, Appl |
| 5 | 69 | 5.1 | 149 | US-10-156-761-8136 | Sequence 8136, Ap |
| 6 | 69 | 5.1 | 178 | US-09-728-626-5680 | Sequence 5680, Ap |
| 7 | 69 | 5.1 | 206 | US-10-156-761-13867 | Sequence 13867, A |
| 8 | 67.5 | 5.0 | 184 | US-09-948-018-8 | Sequence 765, App |
| 9 | 67.5 | 5.0 | 187 | US-10-156-761-8623 | Sequence 8, Appl |
| 10 | 67 | 5.0 | 194 | US-10-127-816-11 | Sequence 8623, Ap |
| 11 | 67 | 5.0 | 202 | US-09-764-868-757 | Sequence 11, Appl |
| 12 | 65 | 4.8 | 179 | US-10-106-698-4858 | Sequence 757, App |
| 13 | 65 | 4.8 | 202 | US-10-189-346-12 | Sequence 4858, Ap |
| 14 | 65 | 4.8 | 202 | US-10-156-761-8978 | Sequence 12, Appl |
| 15 | 64.5 | 4.8 | 191 | US-10-156-761-8978 | Sequence 8978, Ap |

| | | | | | |
|----|------|-----|-----|---------------------|--------------------|
| 16 | 64.5 | 4.8 | 193 | US-09-805-354-8 | Sequence 8, Appl |
| 17 | 64.5 | 4.8 | 193 | US-10-144-259-8 | Sequence 8, Appl |
| 18 | 63.5 | 4.7 | 189 | US-09-815-242-5862 | Sequence 5862, App |
| 19 | 63.5 | 4.7 | 189 | US-09-815-242-12979 | Sequence 12979, A |
| 20 | 63.5 | 4.7 | 202 | US-10-156-761-13460 | Sequence 13460, A |
| 21 | 63.5 | 4.7 | 206 | US-09-738-626-5425 | Sequence 5425, Ap |
| 22 | 62.5 | 4.6 | 144 | US-10-131-406-4 | Sequence 4, Appl |
| 23 | 62.5 | 4.6 | 160 | US-09-882-227-230 | Sequence 230, App |
| 24 | 62.5 | 4.6 | 190 | US-10-156-761-9507 | Sequence 9507, App |
| 25 | 62.5 | 4.6 | 191 | US-09-764-870-390 | Sequence 390, App |
| 26 | 62.5 | 4.6 | 162 | US-10-125-540-390 | Sequence 390, App |
| 27 | 62 | 4.6 | 162 | US-09-738-626-4796 | Sequence 4796, App |
| 28 | 62 | 4.6 | 174 | US-09-864-761-35777 | Sequence 35777, A |
| 29 | 62 | 4.6 | 190 | US-09-738-626-4637 | Sequence 4637, App |
| 30 | 61.5 | 4.6 | 88 | US-10-100-252-6 | Sequence 6, Appl |
| 31 | 61.5 | 4.6 | 142 | US-09-862-027-31 | Sequence 31, Appl |
| 32 | 61.5 | 4.6 | 187 | US-10-136-761-12111 | Sequence 12111, A |
| 33 | 61.5 | 4.6 | 196 | US-09-738-626-6624 | Sequence 6624, Ap |
| 34 | 61.5 | 4.6 | 202 | US-10-189-346-16 | Sequence 16, Appl |
| 35 | 61 | 4.5 | 70 | US-09-864-761-41555 | Sequence 41555, A |
| 36 | 61 | 4.5 | 177 | US-09-791-993-69 | Sequence 69, Appl |
| 37 | 61 | 4.5 | 185 | US-10-101-464A-616 | Sequence 616, App |
| 38 | 61 | 4.5 | 189 | US-09-764-870-523 | Sequence 523, App |
| 39 | 61 | 4.5 | 189 | US-10-125-540-523 | Sequence 523, App |
| 40 | 61 | 4.5 | 189 | US-09-800-729-154 | Sequence 154, App |
| 41 | 60.5 | 4.5 | 203 | US-10-156-761-14346 | Sequence 14346, A |
| 42 | 60.5 | 4.5 | 196 | US-10-219-220-265 | Sequence 265, App |
| 43 | 60 | 4.5 | 134 | US-09-768-235B-40 | Sequence 40, Appl |
| 44 | 60 | 4.5 | 184 | US-09-798-029-4 | Sequence 4, Appl |
| 45 | 60 | 4.5 | 184 | US-09-795-926-46 | Sequence 46, Appl |

ALIGNMENTS

RESULT 1

US-10-102-806-760

Sequence 760, Application US/10102806

Publication No. US2003005442A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103PIC1

CURRENT APPLICATION NUMBER: US/10/102,806

CURRENT FILING DATE: 2002-03-22

PRIOR APPLICATION NUMBER: 09/925,298

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05881

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 846

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 760

LENGTH: 94

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (80)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (91)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-102-806-760

Query Match 34.4%; Score 463; DB 15; Length 94;

Best Local Similarity 98.9%; Pred. No. 2.4e-40;

Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 NDVETSPYTEKHLTDGMYRELCSAATMSDNTANILLTTIGPKELTAFLEHNGDV 134

DB 1 NDVETSPYTEKHLTDGMYRELCSAATMSDNTANILLTTIGPKELTAFLEHNGDV 60

OY 135 TRLDREPELNEAIPNDERDTMPVAMATT 164
 Db 61 TRLDREPELNEAIPNDERDTMPVAMATT 90

RESULT 2

US-10-127-816-9
 ; Sequence 9, Application US/10127816
 ; Publication No. US20030104416A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Klueber, Kevin M.
 ; APPLICANT: Taft, David W.
 ; APPLICANT: Kindvogel, Wayne R.
 ; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
 ; FILE REFERENCE: 01-17
 ; CURRENT APPLICATION NUMBER: US/10/127,816
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/285,408
 ; PRIOR FILING DATE: 2001-04-20
 ; PRIOR APPLICATION NUMBER: US 60/286,482
 ; PRIOR FILING DATE: 2001-04-25
 ; PRIOR APPLICATION NUMBER: US 60/341,050
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 60/341,105
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 09/895,834
 ; PRIOR FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/285,424
 ; PRIOR FILING DATE: 2001-04-20
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-127-816-9

Query Match 5.8%; Score 78; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 6;
 Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEOLGRRIHYSND--LVEYSPYTEKHL-----TDGMTVREL---- 97
 Db 17 LLLAAVLTQTQADVPVPRATRLPEAKDCHIAQFKSLSPKELQAFKAKADIEKRLLEKDL 76
 OY 98 -CSA-----AATMS--DNTAANLLTTIGCPKELTAFLHN 129
 Db 77 RCSSHLFPRANDLKQLOVQERPKALQAEVALTLKVENMTDSALATILGQPLHTLSIHIS 136
 OY 130 MGDHVT-----RLDRMEPELNEAIPNDERDTMPVAMATTLLTGELTTL 176
 Db 137 QLOQTCTQLOATAEPSPSRRLSRMLHRLQEA-QSKETPCGLEASVTSNLFRLTLRLDKCV 195
 OY 177 ASRQOLI 183
 Db 196 ANGDQCV 202

RESULT 3

US-10-142-717-12
 ; Sequence 12, Application US/10142717
 ; Publication No. US20030104579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Immunex Corporation
 ; APPLICANT: Baum, Peter R.
 ; APPLICANT: Mosley, Bruce A.
 ; APPLICANT: Ketchum, Randal R.
 ; APPLICANT: Taylor, Scott L.
 ; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
 ; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
 ; CURRENT FILING DATE: 2002-05-08
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 12
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-10-142-717-12

Query Match 5.8%; Score 78; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 6;
 Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEOLGRRIHYSND--LVEYSPYTEKHL-----TDGMTVREL---- 97
 Db 17 LLLAAVLTQTQADVPVPRATRLPEAKDCHIAQFKSLSPKELQAFKAKADIEKRLLEKDL 76
 OY 98 -CSA-----AATMS--DNTAANLLTTIGCPKELTAFLHN 129
 Db 77 RCSSHLFPRANDLKQLOVQERPKALQAEVALTLKVENMTDSALATILGQPLHTLSIHIS 136
 OY 130 MGDHVT-----RLDRMEPELNEAIPNDERDTMPVAMATTLLTGELTTL 176
 Db 137 QLOQTCTQLOATAEPSPSRRLSRMLHRLQEA-QSKETPCGLEASVTSNLFRLTLRLDKCV 195
 OY 177 ASRQOLI 183
 Db 196 ANGDQCV 202

RESULT 4

US-10-233-926-4
 ; Sequence 4, Application US/10233926
 ; Publication No. US20030131382A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Allen, Stephen M.
 ; APPLICANT: Kinney, Anthony J.
 ; APPLICANT: Falco, S. Carl
 ; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
 ; FILE REFERENCE: BB1419 US NA
 ; CURRENT APPLICATION NUMBER: US/10/233,926
 ; CURRENT FILING DATE: 2002-09-03
 ; PRIOR APPLICATION NUMBER: US/09/735,846
 ; PRIOR FILING DATE: 2000-12-13
 ; NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 4
 ; LENGTH: 149
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; NAME/KEY: UNSURE
 ; LOCATION: (18)
 ; US-10-233-926-4

Query Match 5.2%; Score 69.5; DB 16; Length 149;

Best Local Similarity 20.8%; Pred. No. 30;
 Matches 22; Conservative 25; Mismatches 34; Indels 25; Gaps 4;

OY 156 TMPVAMATTLLKLT-----TGELTTLASROOLIDMMEADRYAGPLRLSALDPAGWFIADK 209
 Db 11 SLSLSLSKSLPLNLAMADHAAEAAPQSSQEEEDMKREAGDGV-----EVDAR 61
 OY 210 SGAGERSRGIIIALGPDGKPSRIIV-----IYTTGSATMDERNR 250
 Db 62 GCGGGAANGCI-----DEGRPIRYVADGIYDLFHRGHAKSLSEQAR 102

RESULT 5

US-10-156-761-8136
 ; Sequence 8136, Application US/10156761
 ; Publication No. US20030119018A1

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8136
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8136
```

```

Query Match      5.1%; Score 69; DB 15; Length 149;
Best Local Similarity 30.9%; Pred. No. 34;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;

QY 168 LTTGELT-----ASRQOLDMENDKVAAGLISALPAGFTADKSG----- 211
Db 50 LLDLTLTKLLKLLVASVDKAKEMGIDWEHD---PALSSHADGRLAEENRLREI 105

QY 212 AGERSGRIIAGLCPDGKPSR 232
Db 106 AGLRDAQALPSAEGPRKPER 126
```

```

RESULT 6
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680
```

```

Query Match      5.1%; Score 69; DB 10; Length 178;
Best Local Similarity 22.1%; Pred. No. 43;
Matches 34; Conservative 19; Mismatches 65; Indels 36; Gaps 5;
```

```

QY 113 LTTTGCPPELTAFLHMGDH--VTRLDREPELNEAIPNDERDTPVMAATLRLKLT 170
Db 27 LKVTAGPKSAVAIKVDSRPLDQLEVEQSQELGFDAAEQGELNFGAGTYLEVSTP 86

QY 171 GELLTLASRQOLDMENDKVAAGLISALPAGFTADKSGAGERSGRIIAGLCPDGK 230
Db 87 G-----VDNPL--TLPRHW-----RRNGRLVALDQDGR 114

QY 231 --SRIVYITGSOATMDERNRQIAETGASLKH 262
Db 115 RVARIGALNDATHTVLIERNKLEVTYTELH 148
```

```

RESULT 7
US-10-156-761-13867
; Sequence 13867, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13867
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13867
```

```

Query Match      5.1%; Score 69; DB 15; Length 206;
Best Local Similarity 23.2%; Pred. No. 53;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;

QY 70 IHYSQNDLVESPTEKHLTDGMTV-----RELCSAATMSDNTAANLLTTI--G 118
Db 1 VHASQC---AOGPETESAGEKVTYDAAQGESFEFVA-----NRSSALLKTAVLLSG 50

QY 119 GPRE-----LTAFLHMGDHTVRLD-----RW-----EPLEALPNDE 152
Db 51 GDRHAEDLLQMLIKADRSRIDPEAYVROVLYIROOVSHRWLRKRRELSVAEP--PE 109

QY 153 RDTTPVMAATLRLKLTGELLTLASRQOLI-----DWMEDKVAAGLISALPAGFT 205
Db 110 ASTGPDAASAELRLVGAALRLARQRTVLYLVTFEDLPEAD--VAIILGCSVTVNS 167

QY 206 IDKSGAGERSGRIIAGLCPDGKPSR 232
Db 168 TTHRSLARLRTLAPELALGPDAAEQEPSPR 197
```

```

RESULT 8
US-10-101-464A-765
; Sequence 765, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
```

```

; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: PRF
; ORGANISM: Pinus radiata
US-10-101-464A-765

Query Match
Best Local Similarity 24.7%; Score 67.5; DB 15; Length 184;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;

QY 41 EPMASTPKVL-----LCGAVLSRIDAGEQGLRIHYSONDIYVSPV-----TEKHLTDG 91
DB 23 FCMTSTERLIVYPYMLGSAVSR-----RDSINGKPAIDWPRKRISLG 67
QY 92 -----MTVRELCSAITYMSDNTAANLL-----TTIGPRELTAFLHMGDHYTRLRW-- 140
DB 68 AARGGLXHHQCDPKRIHROVKANILDEYFEAVYGGGLAKLLDHRSHVTTAVRGTV 127
QY 141 ---EPELINEAIPNDERPTTPVAMATTIRKLITGEL---LTLASRQO--LIDPM 186
DB 128 GHIAPEYLTGSGSSEK--TDVGFGLLELLETITGOKALDFGAAKNGVGLDMV 179

RESULT 9
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: The(11) et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/3/677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-948-018-8

Query Match
Best Local Similarity 22.4%; Score 67.5; DB 10; Length 187;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

QY 112 LLTTTGPKELTAFLHMGDHYTRLDNRPE-----LNEAIPNDERPTTPVAMAT- 163
DB 11 LLLALAAVPPPPAS-----QYCGRLTYWNPDKCCSSCLQRFGP-----PPCGALETG 58
QY 164 -----TLRKLITGELTLASR--COOLDIMMEA-----DKVAGP-----L 195
DB 59 DTWKEASLLPLISRELSSLASQPLSLHLEDELYLEELIYVLDPRPGCGGMAHGTTHIA 118
QY 196 LRSALPAGMTADKSGAGERSGCIITAAIGPDKPSR--IVVITYTGSQATMDERNQIA 253
DB 119 ARYGLIPAAW-----STFAYSLRPSRSPILALEMVAARPPASLSGLQGTFLA 165
QY 254 EIG 256
DB 166 QLG 168
```

```

RESULT 10
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRF
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match
Best Local Similarity 38.6%; Score 67; DB 15; Length 194;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

QY 202 AGMTIADKSGAGERSGSG-----IYALGPDG--KPSRIVITYTT 239
DB 75 SGFRVTPRSGAGERTGAGYRITAIATGPLAVHEPVRVAVVYT 118

RESULT 11
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 202;
```

Best Local Similarity 20.9%; Pred. NO. 84;
Matches , 39; Conservative 26; Mismatches 68; Indels 54; Gaps 7

```

QY 50 LLCAGVSRIDAGQGQRLRIHYSOND--LVBYSPYTEKHL-----TDGMYREL----- 97
Db 17 LLIAVAVLTROTADVPVPATRLTLPVEAKDCHIAQFKSLSPRELOAFKRAKAAIERLLEKDM 76
QY 98 -GSA-----ATMS--DNTAAMILLTTTIGSPKRLTAPLNN 129
Db 77 RCSHLLSRAMDLMQLQVQERPKALQAEVALTLKWEININDSALTTILLOPHTTSHHS 136
QY 130 MGDHYT-----RLDRWEPELNEALPNDERDTPTMVAATTLTKLITGELTTL 176
Db 137 QLOQTQLOQATAEPRPFSRRLSRWHLRLOEA-QSKETPGCELDSTVSNLFQOLLRLKCY 195
QY 177 ASRQOLI 183
Db 196 ASDQCV 202

```

```

RESULT 12
: US-09-764-868-757
: Sequence 757, Application US/09764868
: Patent No. US2002016871A1
:
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT232
: CURRENT APPLICATION NUMBER: US/09/764,868
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - refer to PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1510
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 757
: LENGTH: 179
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-764-868-757

```

| | | | | |
|-----------------------|--------|--------------------|----------------|-------------|
| Query Match | 4.88; | Score 65; | DB 10; | Length 179; |
| Best Local Similarity | 27.38; | Pred. No. 1.1e+02; | | |
| Matches | 24; | Conservative 18; | Mismatches 34; | Indels 12; |
| Gaps | 4 | | | |

| | | | |
|----|-----|---|-----|
| QY | 120 | PKELTAFLHNGDGHVTRID--RMEPELNEA-----IPNDERDTTPVAAATTLRLKLITG | 171 |
| | | : : | |
| | | : : : : : : : : | |
| Db | 52 | PRELTVV---QGEKLEVLVDHSKRRLMWLVKNEAGSGCYIPSNILEPLPGPTGTCQGPSRV | 108 |
| | | : : | |
| | | : : : : : : : : | |
| QY | 172 | ELLTLASR-QQLIDWMEADKRVAGPLLS | 198 |
| | | : : : : : : : : | |
| Db | 109 | PMRLRSSREEVYTDWLAENFSTRATVET | 136 |
| | | : : | |
| | | : : : : : : : : | |

```

RESULT 13
US-10-106-698-4858
; Sequence 4858, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4858
; LENGTH: 179
; TYPE: PRT

```

ORGANISM: Homo sapiens
US-10-106-698-4858

| | | | | |
|-----------------------|-------|-------------------|----------------|-------------|
| Query Match | 4.8% | Score 65; | DB 15; | Length 179; |
| Best Local Similarity | 27.3% | Pred. NO. 11e+02; | | |
| Matches | 24; | Conservative 18; | Mismatches 34; | Indels 12; |
| | | | Gaps | 4 |

| | | | | | |
|----|-----|--|------------------|------------------------|-----|
| QY | 120 | PRELTAETAHNGDHYTRLD--- | RWPEELNEA----- | IPDERDITMPVAMATTLRLRLG | 171 |
| | | : : : | : : : | : : : | : |
| Db | 52 | PRELTVV---QGEKLEIVLDHSEKRMVLVNEARSGYIPSNILEP | LOPGTPGTGOGSPSRV | 108 | |
| | | : : : | : : : | : : : | : |
| QY | 172 | ELLTIAAR-QOLIDMMEADRVAGPIIARS | 198 | | |
| | | : : : : : : : : : : : | : | | |
| Db | 109 | PMLRLSSRPEEVTDLQANENSTAVRT | 136 | | |

```

RESULT 14
US-10-189-346-12
; Sequence 12, Application US/10189346
; Publication No. US20030119162A1
; GENERAL INFORMATION:
; APPLICANT: Churchill, Mafr
; APPLICANT: von Bodman, Susanne
; APPLICANT: Schweizer, Herbert
; APPLICANT: Gould, Ty
; APPLICANT: Hoang, Tung
; APPLICANT: Murphy, Frank
; APPLICANT: Watson, William
; TITLE OF INVENTION: Structural Basis of Quorum Sensing Signal Generation and Metho
; TITLE OF INVENTION: Therapeutic Agents Derived Therefrom
; FILE REFERENCE: 2848-44
; CURRENT APPLICATION NUMBER: US/10/189,346
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: 60/303,449
; PRIOR FILING DATE: 2001-07-04
; PRIOR APPLICATION NUMBER: 60/366,575
; PRIOR FILING DATE: 2002-03-21
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-189-346-12

```

| | Query Match | Best Local Similarity | Score | DB | Length |
|---------|-------------|---|---------------------------|------------|----------|
| Matches | 36; | Conservative | 28.3%; Pred. No. 1,3e+02; | 15; | 202; |
| | | | 18; Mismatches | 41; Indels | 32; Gaps |
| | | | | | 8; |
| QY | 120 | PRLTA-----FLHNGDHYTRLDWE--PELNEALIND--ERTTTPVMAATILRK | 167 | | |
| | 12 | PELAADLDGRYRRRVVEQLG-----WALSANPAFERDPDRDDTVYVMARNAGE | 63 | | |
| Db | 168 | LITGELLITASROOLIDMEADVAG--PLLRSALPAGWTI-----ADKSGAG--ERGS | 217 | | |
| | 64 | MGCARLLPTQPYLTLESIFADLVADVPLPKSA--AVWELSRRAAFADENGPONAEIV | 121 | | |
| QY | 218 | RGIITAL | 224 | | |
| | | :: : | | | |
| QY | 122 | KPMIAIV | 128 | | |

US-10-156-761-8978
RESULT 15
Sequence 8978, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 seconds
(without alignments) 2458.980 Million cell updates/sec

Title: SEQ2_37S

Perfect score: 1345
Sequence: 1 HPELVKVKVDAEDQIGARVG.....TMDERNQIAEIGASLIRKM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 39.0 | 105 | 2 | JC2566 |
| 2 | 78 | 5.8 | 191 | 2 | S67447 |
| 3 | 74 | 5.5 | 177 | 2 | AB3753 |
| 4 | 73.5 | 5.5 | 184 | 2 | T21126 |
| 5 | 72.5 | 5.4 | 200 | 2 | T45195 |
| 6 | 72.5 | 5.3 | 200 | 2 | G97064 |
| 7 | 71.5 | 5.3 | 152 | 2 | T36984 |
| 8 | 71.5 | 5.3 | 195 | 2 | T36975 |
| 9 | 71 | 5.3 | 192 | 2 | G83096 |
| 10 | 70.5 | 5.2 | 131 | 2 | AD2281 |
| 11 | 70 | 5.2 | 108 | 2 | E84217 |
| 12 | 69.5 | 5.2 | 145 | 2 | F84251 |
| 13 | 69.5 | 5.2 | 167 | 2 | D87360 |
| 14 | 69 | 5.1 | 180 | 2 | C71869 |
| 15 | 68.5 | 5.1 | 172 | 2 | AD3606 |
| 16 | 68.5 | 5.1 | 190 | 2 | T09136 |
| 17 | 68 | 5.1 | 170 | 2 | AB0192 |
| 18 | 67.5 | 5.0 | 192 | 2 | AB3587 |
| 19 | 67.5 | 5.0 | 198 | 2 | D95285 |
| 20 | 66.5 | 4.9 | 128 | 2 | F87333 |
| 21 | 65.5 | 4.9 | 42 | 2 | F56978 |
| 22 | 65.5 | 4.9 | 116 | 2 | C82906 |
| 23 | 65.5 | 4.9 | 195 | 2 | AE0623 |
| 24 | 65.5 | 4.9 | 197 | 2 | H90211 |
| 25 | 65 | 4.8 | 148 | 2 | F75283 |
| 26 | 65 | 4.8 | 150 | 2 | AC0284 |
| 27 | 65 | 4.8 | 153 | 2 | E81708 |
| 28 | 65 | 4.8 | 160 | 1 | E69186 |
| 29 | 65 | 4.8 | 168 | 2 | B75498 |

| | | | | | | |
|----|------|-----|-----|---|--------|--------------------|
| 30 | 65 | 4.8 | 180 | 2 | G70912 | hypothetical prote |
| 31 | 65 | 4.8 | 198 | 2 | C69296 | conserved hypotnet |
| 32 | 64.5 | 4.8 | 151 | 2 | D81333 | probable protein-t |
| 33 | 64.5 | 4.8 | 177 | 2 | D90227 | hypothetical prote |
| 34 | 64.5 | 4.8 | 188 | 1 | G69365 | conserved hypotnet |
| 35 | 64 | 4.8 | 113 | 2 | D70580 | hypothetical prote |
| 36 | 64 | 4.8 | 142 | 2 | A71000 | hypothetical prote |
| 37 | 64 | 4.8 | 150 | 2 | T08585 | calmodulin - soybe |
| 38 | 64 | 4.8 | 152 | 2 | AG0769 | probable acetyltra |
| 39 | 64 | 4.8 | 177 | 2 | T48420 | hypothetical prote |
| 40 | 63.5 | 4.7 | 168 | 2 | T20606 | hypothetical prote |
| 41 | 63.5 | 4.7 | 179 | 2 | AB1994 | hypothetical prote |
| 42 | 63.5 | 4.7 | 193 | 2 | S42528 | hypothetical prote |
| 43 | 63.5 | 4.7 | 196 | 2 | H69647 | GTP-binding protei |
| 44 | 63 | 4.7 | 122 | 2 | AC3609 | 2-dehydro-3-deoxy- |
| 45 | 63 | 4.7 | 124 | 2 | F75373 | transcription regu |
| | | | | | | conserved hypotnet |

ALIGNMENTS

RESULT 1
JC2566
bia protein - Pseudomonas aeruginosa (fragment)
C:Species: Pseudomonas aeruginosa
C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996
C/Accession: JC2566
R:West, S.E.H.; Schwelzer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 1481281, 81-86, 1994
A:Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro
A:Reference number: JC2565
A>Note: due to a typographical error the volume number 148 appears as 128
A/Accession: JC2566
A/Molecule type: DNA
A/Residues: 1-105 <MS>
C:Genetics:
A:Gene: bia
C:Superfamily: beta-lactamase I

Query Match 39.0% Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 6.9e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATLRLKLTGELTTLASROQLDMMEDVAVAGLISALPAGWFIADKSGAGGSGK 219
|||||
DB 2 AMATLRLKLTGELTTLASROQLDMMEDVAVAGLISALPAGWFIADKSGAGGSGK 61
QY 220 IIALGPDGKPSRIWVYTTGSGQATMDERNQIAEIGASLIRKM 263
|||||
DB 62 IIALGPDGKPSRIWVYTTGSGQATMDERNQIAEIGASLIRKM 105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996

A:Reference number: 221766

A/Accession: T38062

A>Status: preliminary; translated from GR/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <MC2>

A:Cross-References: EMBL:669944; NID:91217974; PIDN:CAA93808.1; PID:91217978; GSPDB:G

A:Experimental source: strain 972h-; cosmid clF12

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.8%; Score 78; DB 2; Length 191;

Best Local Similarity 22.6%; Pred. No. 31;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;
Db 104 MSDNANLTLTTIGPKELTAFLHMGDVTNRIDRWEPELNEAIPNDRDTPVAMAT 163
1 MSYMSLNLNLNPSGIDKIALV-----VARLD-----PASSKSTAOVSLN 46
QY 164 TLKLLTGLTTLASRQOLDMWEADKVAAPLNSALPAGWFIADKSGAGERSGRTIAA 223
Db 47 EERC-----ILRIPGLIIVNFRKDSPEYMSNAINIGYYE--GLAFLGKQIISI 99
QY 224 LCPDGK-----SRIVYITTSATYDERKQI----- 252
Db 100 ---SRPLEDKMLWSSRFMLDPTLTLYOLLREKTEDEKHQDLASNLASPLCLHWS 155
QY 253 AETGASLTKH 262
Db 156 VENGAGLKH 165

RESULT 3

AB3753
hypothetical protein BH0825 (imported) - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: AB3753
R:Takami, H.; Nakagawa, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: AB3753
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-117 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04544.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0825
C:Superfamily: Bacillus subtilis hypothetical protein ykka

Query Match 5.5%; Score 74; DB 2; Length 177;
Best Local Similarity 20.3%; Pred. No. 60;
Matches 36; Conservative 29; Mismatches 56; Indels 64; Gaps 8;
Db 42 PMSTFVLLCGAVLSRIDAGEOLGRIRHYSNDLVESPYTEKHLTDGMTVRELCSSAA 101
11 PNDTSGVGLTYAVVEVD-----RLH-----HLIEDVPEELTYKG 47
QY 102 ITMSDNTAANLTLTTIGPKELTAFLHMGDVTNRID- RWEPEL-NEAIPND----- 151
Db 48 SDGDENMAQL-----NHLTYVDVAVWVRIRKEALPDSLEAHPM 89
QY 152 -ENDTTPVAMATTLRLKLTGELLTLASRQ-----LIDMWADKVAAGPLLSALP 201
Db 90 VDRKGKLPVYTSLSVQELIERQRYVALLKETCALHDDLARWIPYEERQATIRWGL- 148
QY 202 AGWFIAD 208
Db 149 --WHMAD 153

RESULT 4

T21126
ADP-ribosylation factor homolog F19H8.3 [similarity] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
C:Accession: T21126
R:Steward, C.
submitted to the EMBL Data Library, March 1997
A:Reference number: 219379
A:Accession: T21126
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-184 <WIL>
A:Cross-references: EMBL:T293378; PIDN:CAB07583.1; GSPDB:GN00020; CESP:F19H8.3
A:Experimental source: clone F19H8
C:Genetics:
A:Gene: CESP:F19H8.3
A:Map position: 2
A:introns: 143/2

C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)
F:90-95/Region: nucleotide-binding motif B
F:127-130/Region: GTP-binding NKXD motif
F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

Query Match 5.5%; Score 73.5; DB 2; Length 184;
Best Local Similarity 21.6%; Pred. No. 69;
Matches 46; Conservative 39; Mismatches 79; Indels 49; Gaps 11;
QY 31 ILSEFRSEPRPMSTFVLLCGAVLSRIDAGEOLGRIRHYSNDLVESPYTEKHLTD 90
Db 6 VLKSPK---PSGREIRILTLGL---DNAGKTTILKQL--SSEDVOHVP-----TK 49
QY 91 GMTVRELCSSAATMSDNTAANLTLTTIGPKELTAFLHMGDVTNRID-TRLD 138
Db 50 GFNKTVA-----KGD---IRLVMDIGGORSIRPYMSNYENIDTLIFVIDSNDKRRFD 102
QY 139 RWEPELNEAIPNDRDTPVAMATTLRLKLTGELLTLASRQOLDMWEADKVAAPLRS 198
Db 103 EMNIELGELL-DEEKLRKVPVLIFANKQ-----DLVTAASSEETRLIND-----LLRD 151
QY 199 ALPAGWFIADKSGAERSGRTIAALGPPGKPS 231
Db 152 RT---WHIQACSLANEGINDGITWVASNLKPA 181

RESULT 5

T45195
hypothetical protein u1756t (imported) - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: T45195
A:Accession: T45195
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

Query Match 5.4%; Score 72.5; DB 2; Length 113;
Best Local Similarity 28.6%; Pred. No. 43;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;

QY 130 MGDHY--TRLDWEPELN---EAIIPNDRDTPVAMATTLRLKLTGELLTLASRQ 180
Db 1 MGDGIGMEREGRWTCNTOCPRLRVPGDSEPTLDRASPEDLIIT--NLSLPIHSPPS 58
QY 181 QLIDMNEA-DKVAQPLL-----RSALPAGFIADKSGAGERSGRTIAALGPPGKPSR 232
Db 59 RDDWVEPFDALGTAVFATGDAKMPAVGCI-----GASTRGS-GILASISPFQPAR 112

RESULT 6

G97064
spore coat protein COTC (imported) - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R:Rolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

AD2281
hypothetical protein alr3803 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD2281
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MIMD:21595285; PMID:11759840
A:Accession: AD2281
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-131 <STO>
A:Cross-references: GB:BA000019; PIDN:BAE75502.1; PID:g17132937; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3803

Query Match 5.2%; Score 70.5; DB 2; Length 131;
Best Local Similarity 26.0%; Pred. No. 78;
Matches 34; Conservative 19; Mismatches 51; Indels 27; Gaps 5;
QY 77 LVEYSPVTEKHLTDGKMTVRELCSNA-ITMSDNTANLLTTIGPKRELTAFLHNM--GDH 133
DB 1 MWSSSTVTLQNDINLVLFLEFSSONQOLYDIOPEFLDPYYSKEFNRFITALYAGNF 60
QY 134 VTRLD--RWPELNEAIPNDERDTPYAMATTLKRLT-----GE 172
DB 61 VIFPDMTWMQDENRFVTNPE--LNLADISTIQKLTSVHKERKFCGSLAQMIDNGH 117
QY 173 LFTLASRQOLI 183
DB 118 LFTILQRLQAI 128

RESULT 11
E84217
hypothetical protein Vng0594h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84217
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Laitheuer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jaldic, Jung, K.H.; Alam, M.; Freitas, T. 12176-12181, 2000
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MIMD:20504483; PMID:11016950
A:Accession: E84217
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <STO>
A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA619105.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0594h

Query Match 5.2%; Score 70; DB 2; Length 108;
Best Local Similarity 27.5%; Pred. No. 66;
Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;
QY 44 MSTFEVLLCGAVLSIDAGQDQGRRIHSQNDLVESPVTEKHLTDGKMTVRELCSAAT 103
DB 1 MDVTYTELDGALALRLAARD--RVFEVRFDALEVTDTLRRHGDGDRV-----GSIT 50
QY 104 MSDNTANLLTTIGPKREL-----TAFTHNGDHYTRLDR 139
DB 51 NDDGTDRTMARLTFPDSDPIAVEVPTSEVAALVDAATKTRDR 92

RESULT 12
F84251

hypothetical protein Vng0953c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84251
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Laitheuer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja, Jung, K.H.; Alam, M.; Freitas, T. 12176-12181, 2000
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MIMD:20504483; PMID:11016950
A:Accession: F84251
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <STO>
A:Cross-references: GB:AE004437; NID:g10580510; PIDN:AA619378.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0953c

Query Match 5.2%; Score 69.5; DB 2; Length 145;
Best Local Similarity 20.3%; Pred. No. 11e+02;
Matches 30; Conservative 29; Mismatches 66; Indels 23; Gaps 3;
QY 41 FPMSTFEVLLCGAVLS-----RIDAGQDQGRRIHSQNDLVESPVTEKH 87
DB 3 FSVSGSAALFTAAEFVSGILTSAAFNQFVVDANRARDRVITAKNTVEVANTYDS 62
QY 88 LTDGKMTVRELCSAATMSDNTANLLTTIGPKRELTAFLHNGDHYTRLDRWPELNEA 147
DB 63 VNDITVATNNGSTLSVSQTDVLE--DGEVTDASAYSSVDGNSQTDLM----- 112
QY 148 IPNDERDTPYAMATTLKRLTGLLELT 175
DB 113 LPEGTYSVTVPTGSAPTRVKKVVTGTGVT 140

RESULT 13
D87360
conserved hypothetical protein CC0895 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87360
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, K.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, B.; J.; Ermolav, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MIMD:21173698; PMID:11259647
A:Accession: D87360
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-167 <STO>
A:Cross-references: GB:AE005673; NID:g13422160; PIDN:AAK22880.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0895

Query Match 5.2%; Score 69.5; DB 2; Length 167;
Best Local Similarity 25.4%; Pred. No. 1.3e+02;
Matches 36; Conservative 19; Mismatches 56; Indels 31; Gaps 6;
QY 2 PETLVAKDA-----EDQGARVGYELDIN---SGTLESFSEERFPMASTR 48
DB 24 PWTLLIRDAFAVGRFEQMDRLGVARNVLAARLKLTVSHQVMEYTRYSER-PPRHETW 82
QY 49 VLLCGAVLSIDAGQDQGRRIHSQNDLVESPVTEKHLTDGKMTVRELCSAAT 103
DB 83 LTERGALSPVLLTMAEMGDHYVGR---DKSPVLFRRHKTGCAFPVYLACAGQVYD 138
QY 104 MSDNTANLLTTIGPKRELTA 125
DB 139 RRIIERA-----GPDOLTA 152

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds
(Without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_37S
Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARV...TMDERNRQIAEIGASLIKHW 263

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 78 | 5.8 | 191 | YDA4_SCHPO | Q10346 schizosacch |
| 2 | 69.5 | 5.2 | 196 | RAC2_LORJA | Q40220 lotus japon |
| 3 | 69 | 5.1 | 201 | RACG_DICDI | Q99800 dictyosteli |
| 4 | 68 | 5.1 | 122 | HG3L_THIPE | P60509 thlobacilli |
| 5 | 65.5 | 4.9 | 116 | RBR4_UREPA | Q9PQ10 ureaplasma |
| 6 | 65.5 | 4.9 | 206 | KTHY_METAC | O8THS9 methanosarc |
| 7 | 65 | 4.8 | 198 | OGG1_ARCFU | C98769 archaeospor |
| 8 | 63.5 | 4.7 | 193 | SAR2_LYCES | P52864 lycopersico |
| 9 | 63.5 | 4.7 | 196 | ALKH_BACST | P50846 b.king/kdpy |
| 10 | 63.5 | 4.7 | 200 | RR4_PELNE | O9TSD3 peilla nees |
| 11 | 63 | 4.7 | 176 | HSIV_THEMA | O9WY21 thermotoga |
| 12 | 62.5 | 4.6 | 144 | MANR_THCOLI | P27245 escherichia |
| 13 | 62.5 | 4.6 | 146 | HBB_RABIT | P20209 oryctolagus |
| 14 | 62.5 | 4.6 | 160 | TATB_HELPY | Q25700 helicobacte |
| 15 | 62.5 | 4.6 | 172 | YDEJ_ECOLI | P31131 escherichia |
| 16 | 62.5 | 4.6 | 182 | PYRE_STRCO | Q9X8T7 streptomyce |
| 17 | 62.5 | 4.6 | 184 | HRRP_PSEST | P37929 pseudomonas |
| 18 | 62.5 | 4.6 | 195 | TRPF_THEVO | O979V6 thermoplas |
| 19 | 62 | 4.6 | 173 | NDK_TREPA | O83974 treponema p |
| 20 | 62 | 4.6 | 174 | Y265_BORBU | O51280 borrelia bu |
| 21 | 62 | 4.6 | 174 | Y265_BORBU | O51280 borrelia bu |
| 22 | 62 | 4.6 | 182 | RR4_BELCH | P17940 rattus norv |
| 23 | 62 | 4.6 | 182 | Y861_BIFLO | O89520 bifidobacte |
| 24 | 61.5 | 4.6 | 184 | MLR1_SCHPO | Q9UUG5 schizosacch |
| 25 | 61.5 | 4.6 | 185 | YCJC_ECOLI | P38532 escherichia |
| 26 | 61.5 | 4.6 | 197 | HAN1_XENLA | O73615 xenopus lae |
| 27 | 61 | 4.5 | 177 | VNSC_RINDK | P35948 rinderpest |
| 28 | 61 | 4.5 | 177 | VNSC_RINDK | P35948 rinderpest |
| 29 | 61 | 4.5 | 177 | VNSC_RINDK | P35948 rinderpest |
| 30 | 61 | 4.5 | 194 | ARL2_DROME | O06849 drosophila |
| 31 | 60.5 | 4.5 | 126 | YGM1_YEAST | O20234 iris ensata |
| 32 | 60.5 | 4.5 | 142 | VR63_METTM | P53130 saccharomyc |
| 33 | 60.5 | 4.5 | 150 | PDUV_SALTY | O50770 methanobact |
| | | | | | Q9XDM6 salmoneella |

| | | | | | | |
|----|------|-----|-----|---|------------|--------------------|
| 34 | 60.5 | 4.5 | 152 | 1 | YU33_YERPE | O8XCF8 yersinia pe |
| 35 | 60.5 | 4.5 | 160 | 1 | TATB_HELPJ | O9ZM58 helicobacte |
| 36 | 60.5 | 4.5 | 178 | 1 | HSIV_RALSO | O8Y347 ralsionia s |
| 37 | 60 | 4.5 | 121 | 1 | SECR_HUMAN | P09683 homo sapien |
| 38 | 60 | 4.5 | 161 | 1 | PIN_BPR4 | P07068 bacterioph |
| 39 | 60 | 4.5 | 178 | 1 | PIRE_ARCFU | O28533 archaeoglob |
| 40 | 60 | 4.5 | 178 | 1 | UCRI_ANASP | P70758 anabaena sp |
| 41 | 60 | 4.5 | 200 | 1 | TATB_CAUCR | Q9A6T1 caulobacter |
| 42 | 60 | 4.5 | 203 | 1 | HOP2_YEAST | P53187 saccharomyc |
| 43 | 59.5 | 4.4 | 103 | 1 | RS10_NEIGO | P48851 neisseria g |
| 44 | 59.5 | 4.4 | 156 | 1 | BFR_AZOVI | P22759 azotobacter |
| 45 | 59.5 | 4.4 | 177 | 1 | ATPD_HAEIN | P43717 haemophilus |

ALIGNMENTS

RESULT 1
YDA4_SCHPO STANDARD: PRT: 191 AA.
ID YDA4_SCHPO
AC Q10346;
DE 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DI 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1P12.04C in chromosome I.
GN SPAC1P12.04C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_Taxid:4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:972;
RX MEDLINE-21848401; PubMed-11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather E.J., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grympep B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Meest D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann M., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.,
RT The genome sequence of Schizosaccharomyces pombe.
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: Z69944; CAA93808.1; -
DR PIR: T38062; S67447.
DR GeneDB: Spombe; SPAC1P12.04C; -

QY 44 MSTRVVLGCAVLISRIDAGEOJGRTHYSONDLY-EXSP-VTEKHLTDGMTVRELCSAA 101
 DB 1 MKSIVCVVG-----EGGIGKTSMLSYTSNSISNEYQPYFDFKYST----- 42
 QY 102 ITMSDNTANLLTTTIGSPRELTA-----LHNKGDAVTRLDREPELN 145
 DB 43 LHMNKKPYNLMLMTAGQEEFSKRLRLSPQNDVFLTCFSLINPSSESNILDSWQELN 102
 QY 146 EAPNDRDTPMPVAMATTLRLKLTGEL--TLASROOL 182
 DB 103 ENCPN-----TPIVLVGMOLDKSNVILRLCEKROL 135

RESULT 4

HC3L_THIFE STANDARD; PRT; 122 AA.

AC P80509;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cytochrome-C3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
 DE (Fragments).
 GN HOG.
 OS Rhodococcus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 CX NCBI_TaxID=920;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=ATCC 19859;
 RX MEDLINE=96241862; PubMed=8661919;
 RA Fischer J., Quentmeyer A., Kostka S., Kraft R., Friedrich C.G.;
 RT "Purification and characterization of the hydrogenase from
 RT Thiodiobacillus ferrooxidans."
 RU Arch. Microbiol. 165:289-296(1996).
 CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 - 2 H(+) + 2
 CC ferriocytochrome c3.
 CC -1- COFACTOR: IRON.
 CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
 CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
 CC KM Oxidoreductase; Iron.
 FT NON_CONS 20 21
 FT NON_CONS 29 30
 FT NON_CONS 35 36
 FT NON_CONS 42 43
 FT NON_CONS 59 60
 FT NON_CONS 72 73
 FT NON_CONS 78 79
 FT NON_CONS 87 88
 FT NON_CONS 98 99
 FT NON_CONS 107 108
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 13146 MW; EBAFAF365348DA0B CRC64;

Query Match 5.1%; Score 68; DB 1; Length 122;
 Best Local Similarity 26.4%; Pred. No. 73;
 Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;

QY 19 VGYIELDNLSKILSFSEERFPMSTFKVLLCGAVLSRIDAGEOJGRTHYSONDLY 78
 DB 12 VGRREGSDLSKILEYFRN-----ALLAFGGG-----LG 41
 QY 79 EYSPVTEKHLTDGMTVRELCSAAITMSDNTANLLTTTIGSPREL 124
 DB 42 KYAPFTGTNYEIGVT-----ISGDKRDPQAGLVYT-----PREST 75

RESULT 5

REBA_UREPA STANDARD; PRT; 116 AA.

ID RBFA_UREPA
 AC O9POH0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ribosome-binding factor A.

GN RBFA OR U0321.

OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.

OX NCBI_TaxID=134821;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Serovar 3;

RX MEDLINE=20500219; PubMed=11048724;

RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,

RA Cassell G.H.;

RT "The complete sequence of the mucosal pathogen Ureaplasma

urealyticum."

RL Nature 407:757-762(2000).

CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not

with 30S subunits that are part of 70S ribosomes or polysomes).

CC Essential for efficient processing of 16S rRNA. May interact with

the 5'terminal helix region of 16S rRNA (by similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).

CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; AE002129; AAF30730.1; -

DR HAMAP; MF_000003; -1.

DR InterPro: IPR000238; Rib_bind_facta.

DR Pfam; PF02033; RBFA; 1.

DR ProDom: PD007327; Rib_bind_facta; 1.

DR TIGRPFAMs; TIGR00082; rbfA; 1.

DR PROSITE; PS01319; RBFA; FALSE_NEG.

KW rRNA processing; Complete proteome.

SQ SEQUENCE 116 AA; 13247 MW; A79DCG71F0547514 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 116;
 Best Local Similarity 28.9%; Pred. No. 1,le+02;
 Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

QY 144 LNEAIPNDRDTPMPVAMATTLRLKLTGEL-----TLASROOLDMMF-ADYVAGEL 196
 DB 18 INNALANEINDKIALIARTAYR--LSNDLSVAKIFLDHAKRSMKLVLENVNVSG-L 74
 QY 197 RSALPAGW-----FIADKS 210
 DB 75 RSKLAEMTSYKVPRLRFVIDET 97

RESULT 6

KTHY_METAC STANDARD; PRT; 206 AA.

ID KTHY_METAC
 AC O8THS9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN TKR OR MA4433.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2214;

RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels S., Smirnov S., Ancoor D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,

RA Linton L., McEwen P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers R.R., Swanson R.V., Zinder S.H., Zander E.,
RA Metcalf W.W., Birren B.,
RA "The genome of *Methanocaldococcus jannaschii* reveals extensive metabolic
RT and physiological diversity."*,
RL Genome Res. 12:532-542(2002).
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate - ADP + thymidine
CC 5'-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE011164; AAM07774.1; -
DR HAMAP: MF_00165; 1.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR TIGRFAMs: TIGR00041; DTM_Kinase; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE NEG.
KW Transferase; kinase; Nucleotide biosynthesis; ATP-binding;
KM Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 23303 MW; A19C25B5C85423B3 CRC64;

Query Match 4.98; Score 65.5; DB 1; Length 206;
Best Local Similarity 18.78; Pred. No. 2.2e+02;
Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;

QY 56 LSRID-AGOEQLGRIRHNSDLYEVSPTKHLTDGTVRELCSNAITMSDNTAAILLL 114
DB 8 LKIDSSGSKSTVAKKIQ-ENSELAFVEFVETREPRG-----TLTGAVENAIQ 55
QY 115 TTIGGKRELTAFLHNGDHTVTRLDWEPELNEAIPNDERDTTPVMAATLTKLLTGEHL 174
DB 56 SPIDQLAEFLFADHAEHLAKLVKPALEDKTVISDRSVAYOGITLKRNLNPL- 114
QY 175 TLASROQLIDWMEADKVAAGPLRLSALPAGW-----FIAD-----KSG-AGE- 214
DB 115 -----EMVR-----DLHGMVIVIPDLTFLFDIEPEIAVKRCGRGQTKF 154
QY 215 -----RGSRGITAAIGPDGKPSRIVVYTTGSGQATMD 247
DB 155 EKIEFLRGVRELFGLALAE-EPERFVIVDASSGPEDEK 192

RESULT 7
OGGI_ARCFU STANDARD; PRT; 198 AA.
ID OGGI_ARCFU
AC 029876;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable N-glycosylase/DNA lyase [includes: 8-oxoguanine DNA
DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
DE (EC 4.2.99.18) (AP lyase)].
GN OGC OR AF0371.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_Taxid=2234;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;
RA Klank H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwin M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyprides N.C.,
RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodok A., Zhou L.,
RA Overbeek R., Gocayne J.D., Meldrum J.F., McDonald L., Uutterback T.,
RA Cotton M.D., Spriggs T., Arlanch P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.V., Fraser C.M., Smith R.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon *Archaeoglobus fulgidus*."*,
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
CC guanine (7,8-dihydro-8-oxoguanine - 7-oxoG) from DNA. Also nicks
CC DNA at apurinic/apyrimidinic sites (AP sites) (By similarity).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001079; AAB90876.1; -
DR PIR: C69296; C69296.
DR TIGR: AF0371; -
DR HAMAP: MF_00241; -; 1.
DR InterPro: IPR003265; Bndc_3c.
DR Pfam: PF00730; Hnh-GPD; 1.
DR SMART: SM00478; ENDO3c; 1.
KW Hydrolyase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
KW Multifunctional enzyme; Complete proteome.
FT ACT_SITE 122 122
SQ SEQUENCE 198 AA; 22639 MW; 3A5C03AA12F3FFB CRC64;

Query Match 4.88; Score 65; DB 1; Length 198;
Best Local Similarity 24.68; Pred. No. 2.3e+02;
Matches 51; Conservative 25; Mismatches 85; Indels 46; Gaps 14;

QY 14 QGARYGYELDLN---SGKILSFSEERPEPMSTFVLLCGAVLSRIDAQO-----EQ 65
DB 15 QGEEK-GEVEFDRPFLDFSVKATIRTELARCISTANSATAGLKFQRL-LGGVGKKA 72
QY 66 L---GRRIRHSQNDLYEVSPTKHLTDGTVRELCSNAITMSDNTAAILTTTG-GR 121
DB 73 LTLGVRFH---NRKAEY-----IREAFKFLVERALKESSKARILLIKIKGLMK 122
QY 122 ELTAFILHNG-DIVTRLD-----RWEPELNEAIPNDERDTTPVMAATLTKLL-----T 170
DB 123 EASHFLRNGREVVAIIDRIHLRLWLEROGYEPG-----TWAKRYLVEKILMEISEER 177
QY 171 GELLTLASROQLIDWMEADKVAAGPLRLSALPAGW-----FIAD-----KSG-AGE- 214
DB 178 GESL---AEMLDLRIWAE---MTGKVLK 198

RESULT 8
SAR2_LYCES STANDARD; PRT; 193 AA.
ID SAR2_LYCES
AC P52884;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE GTP-binding protein SAR2.

GN SAR2.
 OC Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_Taxid=4081;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pericarp;
 RX MEDLINE=94169306; PubMed=8123794;
 RA Davies C.;
 RT *Cloning and characterization of a tomato GTPase-like gene related to
 RT yeast and Arabidopsis genes involved in vesicular transport.;
 RL Plant Mol. Biol. 24:525-531(1994).
 CC -1 FUNCTION: INVOLVED IN TRANSPORT FROM THE ENDOPLASMIC RETICULUM
 CC TO THE GOLGI APPARATUS (BY SIMILARITY).
 CC -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. SAR1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L12051; AAA34168.1; -
 DR PIR: P32528; S42528.
 DR HSSP: P32889; 1RRG.
 DR InterPro: IPR006689; ARF/SAR.
 DR InterPro: IPR001806; Ras_trnsfrmng.
 DR InterPro: IPR006687; SAR1.
 DR InterPro: IPR005225; small_GTP.
 DR Pfam: PF000025; arf.1.
 DR PRINTS: PR00449; RASTRNSFRMNG.
 DR PRINTS: PR00328; SAR1GTPBP.
 DR SMART: SM00178; SAR.1.
 DR TIGRFAMs: TIGR00231; small_GTP.1.
 DR PROSITE: PS01020; SAR1.1.
 DR GTP-binding; Protein transport; Endoplasmic reticulum; Golgi stack.
 FT NP_BIND 27 34 GTP (BY SIMILARITY).
 FT NP_BIND 70 73 GTP (BY SIMILARITY).
 FT NP_BIND 129 132 GTP (BY SIMILARITY).
 SQ SEQUENCE 193 AA; 21923 MW; CFS223DB9DE9984F CRC64;
 Query Match 4.7%; Score 63.5; DB 1; Length 193;
 Best Local Similarity 22.5%; Pred. No. 3e+02;
 Matches 42; Conservative 35; Mismatches 61; Indels 49; Gaps 12;
 OY 14 QLCARVGYIELDLSNGK--ILESFRSEE-----RFPKMSFKVLLCGAV-LSRIDAGO 63
 DB 18 QXAKKILFLID-NAGKTTLLHMLKDEKLVQHQPQYV--TSEELSTGINKKRAFDLGG 73
 OY 64 EQLGRRI--HYSDNDLVEY-----SPYTERKHLFDGMTVRELCSAITMSDNAN 111
 DB 74 HQIARVWMDYVAVDAVYLVYDANDRERPEAKREL-DGL-----LSDESLTN 121
 OY 112 ILITTTGGPKREL-----TAFLLNMGDHYTRLDREPELNEAIPNDERDTTTPVAMATTL 165
 DB 122 VPELLIGKMKIDIPYASEDELRYHLG--LTGVYTGKGNINLNGTNR-----PIEYFMCSTIV 176
 OY 166 RKLITGE 172
 DB 177 RKMGYGE 183
 RESULT 9
 ALKH_BACSU STANDARD; PRT; 196 AA.
 AC P50846;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE KHC/KDPG aldolase [includes: 4-hydroxy-2-oxoglutarate aldolase
 DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KHC-aldolase); 2-
 DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
 DE dehydro-3-deoxygluconate aldolase) (Phospho-2-keto-3-deoxygluconate
 DE aldolase)] (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
 DE aldolase)).
 GN KDCG.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_Taxid=1423;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Marburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehlich S.D.,
 RA Serror P.;
 RT *Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sera and kdg loci cloned in a yeast artificial chromosome.;
 RL Microbiology 142:2005-2016(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bortero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borisov R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaber-Bianchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina M., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priescan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Solido B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpiltra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
 RT *The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis..
 RL Nature 390:249-256(1997).
 CC -1 CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate = pyruvate +
 CC glyoxylate.
 CC -1 CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate +
 CC pyruvate + D-glyceraldehyde 3-phosphate.
 CC -1 PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
 CC -1 PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 CC OF GLYOXYLATE.
 CC -1 SUBUNIT: Homotrimer (By similarity).
 CC -1 SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1 SIMILARITY: BELONGS TO THE KHC/KDPG ALDOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

```

DR EMBL; L47838; AAB38480.1; -.
DR EMBL; 299115; CAB14127.1; -.
DR PIR; H69647; H69647.
DR Subtilist; BG11396; kda.
DR InterPro; IPR000887; Aldose_KDPG_KMG.
DR Pfam; PF01081; Aldolase; 1.
DR TIGRfam; TIGR01182; eda; 1.
DR PROSITE; PS00159; ALDOLASE_KDPG_KMG_1; FALSE_NEG.
DR PROSITE; PS00160; ALDOLASE_KDPG_KMG_2; 1.
DR Lysase; Schiff base; Multifunctional enzyme; Complete proteome.
FT ACT_SITE 43 43 BY SIMILARITY.
FT BINDING 130 130 SCHED-BASE WITH KMG OR PYRUVATE (BY
FT SEQUENCE 196 AA; 20865 MW; EB0B597FD1102CC4 CRC64;
SQ
Query Match 4.7%; Score 63.5; DB 1; Length 196;
Best Local Similarity 22.1%; Pred. No. 3.1e+02;
Matches 27; Conservative 17; Mismatches 59; Indels 19; Gaps 4;
OY 31 ILSEFRSEERFPMSTFKVLLCGAVLSRIDAGQ-EQLGRRIHYSON--DIVESPVTEK 86
DB 54 IIESRRNRDI-----LIGACTVISAOAGGAAGKAGQFIVSGFADLAHLSPVKT 106
OY 87 HLTGNTVRELCSAATMSDNTAANT-----LTTGGPKELTAFLHNMGDHVTILD 138
DB 107 HTPGVLPSEIMELTFCTTLKLPFGVGFIPFMKMLAGPQVTFPIPTGIGHPSEVP 166
OY 139 RW 140
DB 167 DW 168
RESULT 10
RR4_PELNE STANDARD; PRT: 200 AA.
ID RR4_PELNE
AC Q9FSD9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Pella neesiana (Liverwort).
OC Chlorophyta.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Jungermanniopsida; Metzgeriidae; Fossombroniales; Pellineae;
OC Pellineae; Pella.
OC NCBI_TaxID=70144;
OX (1)
RN RP SEQUENCE FROM N.A.
RC TISSUE=Gametophyte;
RA Capeus I., Biocher R.;
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: One of the primary rRNA binding proteins. It binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (by similarity).
CC -1- FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (by similarity).
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (by similarity).
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: BELONGS TO THE S4 FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC EMBL; AJ250456; CAC14050.1; -.
DR HSSP; P81288; IC05.
DR HAMAP; MF_01306; -. 1.
DR InterPro; IPR001912; Ribosomal_S4.
DR InterPro; IPR002942; S4.
DR InterPro; IPR005709; S4_bact.org.
DR Pfam; PF00163; Ribosomal_S4; 1.
DR Pfam; PF01479; S4; 1.
DR SMART; SM00363; S4; 1.
DR TIGRfam; TIGR01017; rpsD_bact; 1.
DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
DR PROSITE; PS50889; S4; 1.
DR Ribosomal protein; rRNA-binding; rRNA-binding; chloroplast.
FT DOMAIN 90 150 S4 RNA-BINDING.
FT SEQUENCE 200 AA; 23178 MW; 0DD0098CD9D933F5 CRC64;
SQ
Query Match 4.7%; Score 63.5; DB 1; Length 200;
Best Local Similarity 22.8%; Pred. No. 3.1e+02;
Matches 39; Conservative 31; Mismatches 78; Indels 23; Gaps 7;
OY 14 QLGRVGYE--LDLNSKILSFSEERFPMSTFKVLLCGAVLSRIDAGQEQGLRRH 71
DB 14 RLGLALPGLTSKILESGYIGSTPNKR-----VSQYRIRL-----EEKQLRFH 58
OY 72 Y--SONDLVERSPVTEKHLTDGMYRELCSAATMSDNTAANT-LTTGGPKELTAFLH 128
DB 59 YGLTEROLKLYRIK--AKSGTGQILSOTLEMRLDNIIFRLGMSPTIGARLVNRRH 116
OY 129 NM-GDHYTRIDRWEPELNEAIPNDERDTTPVAMATTLRLNGELTLAS 178
DB 117 ILINDNYDIPSYNCEPRDYITVNNKESYIINMDSRKPYPNHTFPDS 167
RESULT 11
HSLV_THEME STANDARD; PRT: 176 AA.
ID HSLV_THEME
AC Q9WY21;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease half (EC 3.4.25.-).
GN HSLV OR TM0521.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=23336;
OX (1)
RN RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.R., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT *Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.*;
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (by similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```


| DB | 105 | AAICGQCHQLVGGDDHLQELTKNLTADENVATLEYLLKRVLP | 144 |
|------------|--|--|--------------|
| RESULT 13 | | | |
| HBG_RABIT | | | |
| ID | HBG_RABIT | STANDARD: | PRT: 146 AA. |
| AC | P02099: | | |
| DT | 21-JUL-1986 (Rel. 01, Created) | | |
| DT | 21-JUL-1986 (Rel. 01, Last sequence update) | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | |
| DE | Hemoglobin gamma chain (beta-3). | | |
| HBG: | | | |
| OS | Oryctolagus cuniculus (Rabbit). | | |
| OC | Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus. | | |
| OX | NCBI_TaxID=9986; | | |
| RN | | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=82053017; PubMed=6271761; | | |
| RA | Hardison R.C.; | | |
| RL | *The nucleotide sequence of rabbit embryonic globin gene beta 3.*; | | |
| RL | J. Biol. Chem. 256:11780-11786(1981). | | |
| RN | (2) | | |
| RP | SEQUENCE FROM N.A. | | |
| RX | MEDLINE=89178632; PubMed=2486295; | | |
| RA | Margot J.B., Demers G.W., Hardison R.C.; | | |
| RT | *Complete nucleotide sequence of the rabbit beta-like globin gene | | |
| RT | cluster. Analysis of intergenic sequences and comparison with the | | |
| RT | human beta-like globin gene cluster.*; | | |
| RL | J. Mol. Biol. 205:15-40(1989). | | |
| CC | -1- FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE | | |
| CC | GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE | | |
| CC | HUMAN GAMMA CHAIN GENE, WHICH CODES FOR A FETAL GLOBIN. | | |
| CC | -1- SUBUNIT: Heterotetramer of two alpha chains and two gamma chains. | | |
| CC | -1- TISSUE SPECIFICITY: Red blood cells. | | |
| CC | -1- SIMILARITY: Belongs to the globin family. | | |
| CC | ----- | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | |
| CC | use by non-profit institutions as long as its content is in no way | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | |
| CC | entities requires a license agreement (See http://www.lsb.ch/announce/ | | |
| CC | or send an email to license@lsb.ch). | | |
| CC | ----- | | |
| DR | EMBL: M18818; AAA02984.1; - | | |
| DR | EMBL: V00883; CAA24252.1; - | | |
| DR | PIR: A02417; HRR3. | | |
| DR | HSSP: P02100; IAW3. | | |
| DR | InterPro: IPR002337; Beta_haem. | | |
| DR | InterPro: IPR000971; Globin. | | |
| DR | Pfam: PF00042; Globin. 1. | | |
| DR | PRINTS: PRO0814; BEPAAEM. | | |
| DR | PROSITE: PS01033; GLOBIN; 1. | | |
| KW | Heme; Oxygen transport; Transport; Erythrocyte; Embryo. | | |
| KW | FT INIT MET 0 0 | | |
| SO | SEQUENCE 146 AA; 16093 MW; EBB8D6C1C24DD2D82 CRC64; | | |
| QY | Query Match 4.6%; Score 62.5; DB 1; Length 146; | | |
| QY | Best Local Similarity 30.8%; Pred. No. 2.5e+02; | | |
| QY | Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6 | | |
| DB | 3 FTAEKKAIVTSTWK-----LVDDVEDAGEALGRLT-----LVVY--PWTOREFDSFGNL 48 | | |
| QY | 35 FRSERRPMMSTFVLLCGAVLSRIDAGCEQGRIRHISQNDIVESPVTEKHILTDGTV 94 | | |
| QY | 95 RELCSAIVTMSDNTAANLLITITGCKELTAFLHNGGDHYTRL 138 | | |
| DB | 49 SS--SSAIMGNPVKYKAH-----GKKVLTAF---GDAVKND 79 | | |
| RESULT 14 | | | |
| TATB_HELPY | | | |

```

ID  TATF_HELPY      STANDARD:      PRT:      160 AA.
DC  025700:
DT  16-OCT-2001 (rel. 40, Created)
DT  16-OCT-2001 (rel. 40, Last sequence update)
DT  28-FEB-2003 (rel. 41, Last annotation update)
DE  Sec-independent protein translocase protein tatB homolog.
CN  TATB OR Hsp1060.
OS  Helicobacter pylori (Campylobacter pylori).
OC  Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC  Helicobacteriaceae; Helicobacter.
RX  NCBI_TaxID=210;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=26635 / ATCC 700392;
RX  MEDLINE=97394467; PubMed=9252185;
RA  Tomb J.-F., White O., Kurlavage A.R., Clayton R.A., Sutton G.G.,
RA  Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA  Nelson K., Quackenbush J., Zhou L., Kirnesh E.F., Peterson S.,
RA  Loftus B., Richardson D., Dodson R., Khakh H.G., Glodok A.,
RA  McKenney K., Flitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA  Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA  Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watney L., Wallin E.,
RA  Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA  Venter J.C.;
RT  The complete genome sequence of the gastric pathogen Helicobacter
RT  pylori.
RL  Nature 388:539-547(1997).
CC  -1- FUNCTION: Required for correct localization of precursor proteins
CC  bearing signal peptides with the twin arginine conserved motif
CC  S/T-R-R-X-F-L-R. This sec-independent pathway is termed TAT for
CC  twin-arginine translocation system. This system mainly transports
CC  proteins with bound cofactors that require folding prior to export
CC  (By similarity).
CC  -1- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC  -1- SIMILARITY: Belongs to the tatB family.
CC  CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC  at the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  CC
DR  EMBL; AE000613; ABD08110.1; -.
DR  PIR; D64652; D64652.
DR  TIGR; HPI060; -.
DR  HAMAP; MF_00237; -. 1.
DR  InterPro; IPR003369; Mta_Hcf106.
DR  InterPro; IPR003998; TatB.
DR  Pfam; PF02416; Mta_Hcf106; 1.
DR  PRINTS; PR01506; TATBP/PROTEIN.
DR  TIGRFAMs; TIGR01410; tatB; 1.
KM  Transpore; Protein transport; Translocation; Transmembrane;
KW  Inner membrane; Complete proteome.
FT  TRANSMEM 1
FT  SEQUENCE 160 AA; 18332 MW; 3BD44234F7BABBD7 CRC64;
SQ  POTENTIAL.
Query Match 4.6%; Score 62.5; DB 1; Length 160;
Best Local Similarity 20.5%; Pred. No. 2.8e+02;
Matches 27; Conservative 30; Mismatches 64; Indels 11; Gaps 5
OY 35 FRSEERPFPMGSETFKILGCAVLISDAGOEQLRIRHSQ--NDLVESPYTEKHITD-- 90
DB 19 PLGPEEQAYVDVYKFFRAVAKKTLDAKDTLLKEINIEIKETLEYOKLFENKVESLK 78
OY 91 GATRELCSSAATMSDNTAANE--LTTTGGPKELFAFLNMGDVTRLDMREBELNAEI 148
DB 79 GVKIIEELDAKYT-ABNEIKSIODLMODYOKSLFTNTPRHNLDEYVN----EALNREV 133
OY 149 PNDERDTMPVA 160
DB 134 SDESDEPREVOIA 145

```

```

RESULT 15
ID YDEJ_ECOLI STANDARD; PRT; 172 AA.
AC P31131:
DT 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Hypothetical protein ydeJ.
GN ydeJ OR B1537.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93186717; PubMed=8383113;
RA Cohen S.P., Haechler H., Levy S.B.;
RT Genetic and functional analysis of the multiple antibiotic
RL resistance (mar) locus in Escherichia coli.;
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT The complete genome sequence of Escherichia coli K-12.;
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RA MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiiuchi T.;
RT A 570-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 28.0-40.1 min region on the linkage map.;
RN DNA Res. 3:363-377(1996).
CC -1- SIMILARITY: BELONGS TO THE CINA FAMILY. STRONG, TO E.COLI YGAD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96235; -; NOT ANNOTATED_CDS.
DR EMBL; AE000251; AAC74610.1; -
DR EMBL; D90796; BAA15227.1; -
DR EMBL; D90797; BAA15239.1; -
DR PIR; D64908; D64908.
DR EcoGene; Egl1645; ydeJ.
DR HAMAP; MF_00226; atypical; 1.
DR InterPro; IPR003396; CINA.
DR Pfam; PF02464; CINA; 1.
DR TIGRfam; TIGR00199; cina_c1term; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 172 AA; 16321 MW; 700A7A95C967C868 CRC64;
Query Match 4.6%; Score 62.5; DB 1; Length 172;
Best Local Similarity 27.1%; Pred. No. 3.1e+02;
Matches 23; Conservative 9; Mismatches 26; Indels 27; Gaps 3;

```

```

OY 73 SONDLEVSPTKEHLTDGMTVELCSAATNSDNTAANLLFTTIGSP----- 120
DB 79 SOOSLEFRYSAVSEK-----VAAEKATGAIERAD-ADVSIATITGYGSGPEGEGDGTPTAGTV 131
OY 121 -----KEULTAFIHNMGDHYTRL 137
DB 132 WFAWHIKGQNTYAVVHFAFGDCEFTVL 156

```

Search completed: September 10, 2003, 12:20:49
 Job time : 5.57143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_37S

Perfect score: 1345
Sequence: 1 HPETLVKVDADQLGARVG.....TMDERNRQIAIEGASLIKRW 263

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 39.0 | 105 | 2 | 052026 |
| 2 | 524 | 39.0 | 105 | 2 | 052330 |
| 3 | 493 | 36.7 | 145 | 2 | 08RTD8 |
| 4 | 488 | 36.3 | 102 | 2 | 052639 |
| 5 | 458 | 34.1 | 138 | 2 | 0990F3 |
| 6 | 458 | 34.1 | 138 | 2 | 09AMW1 |
| 7 | 458 | 34.1 | 139 | 2 | 09AMA2 |
| 8 | 456 | 33.9 | 138 | 2 | 09AMAO |
| 9 | 456 | 33.9 | 138 | 2 | 09AM99 |
| 10 | 456 | 33.9 | 139 | 2 | 09AM98 |
| 11 | 453 | 33.7 | 139 | 2 | 09AMW3 |
| 12 | 338 | 25.1 | 67 | 2 | 053553 |
| 13 | 320 | 23.8 | 95 | 2 | 08GDE5 |
| 14 | 316 | 23.5 | 62 | 2 | 09JN58 |
| 15 | 299 | 22.2 | 128 | 2 | 08VQ00 |
| 16 | 245 | 18.2 | 180 | 2 | 08KVT2 |

| | | | | | | |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 242 | 18.0 | 52 | 2 | 09R412 | 09r412 shigella fl |
| 18 | 241 | 17.9 | 48 | 2 | 09RLH0 | 09rlh0 proteus mtr |
| 19 | 197 | 14.6 | 38 | 2 | P97145 | P97145 escherichia |
| 20 | 173.5 | 12.9 | 134 | 2 | 08VTL3 | 08vul3 staphylococ |
| 21 | 128.5 | 9.6 | 109 | 2 | 053698 | 053698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | 053698 | P97146 escherichia |
| 23 | 87 | 6.5 | 100 | 2 | 093505 | 093505 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | 09XB22 | 09xb2 bacillus ce |
| 25 | 79 | 5.9 | 198 | 2 | 09ACM8 | 09acm8 streptococc |
| 26 | 78 | 5.8 | 153 | 5 | 08SSZ0 | 08ssz0 clona intes |
| 27 | 77 | 5.7 | 181 | 16 | 08D108 | 08d108 yersinia pe |
| 28 | 76.5 | 5.7 | 202 | 16 | 098520 | 098520 rhizobium l |
| 29 | 74 | 5.5 | 177 | 16 | 09KEW7 | 09ken7 bacillus ha |
| 30 | 73.5 | 5.5 | 184 | 5 | 045379 | 045379 caenorhabdi |
| 31 | 73 | 5.4 | 175 | 17 | 08RTH8 | 08rtl8 pyrobaculum |
| 32 | 73 | 5.4 | 201 | 17 | 08RM84 | 08rm84 methanocarl |
| 33 | 72.5 | 5.4 | 113 | 2 | 049970 | 049970 mycobacteri |
| 34 | 72.5 | 5.4 | 145 | 2 | 005984 | 005984 staphylococ |
| 35 | 72.5 | 5.4 | 200 | 16 | 09JJE8 | 09jje8 clostridium |
| 36 | 71.5 | 5.3 | 152 | 16 | 09RI91 | 09ria0 streptomyc |
| 37 | 71.5 | 5.3 | 195 | 16 | 09RIA0 | 09x910 yersinia en |
| 38 | 71 | 5.3 | 131 | 2 | 09X9H0 | 09x910 yersinia en |
| 39 | 71 | 5.3 | 169 | 17 | 0972W2 | 0972w2 sulfolobus |
| 40 | 71 | 5.3 | 192 | 16 | 09HW07 | 09hw07 pseudomonas |
| 41 | 71 | 5.3 | 203 | 4 | 096B23 | 096b23 homo sapien |
| 42 | 70.5 | 5.2 | 131 | 16 | 08YOL8 | 08yol8 anabaena sp |
| 43 | 70 | 5.2 | 108 | 17 | 09HRO3 | 09hrg3 halobacteri |
| 44 | 70 | 5.2 | 150 | 10 | 094IG4 | 094ig4 nicotiana t |
| 45 | 70 | 5.2 | 205 | 6 | 09N275 | 09n275 ovis aries |

ALIGNMENTS

RESULT 1

052026 AC 052026 PRELIMINARY; PRT; 105 AA.

DT 01-NOV-1996 (TRENBLUREL. 01, Created)

DT 01-NOV-1996 (TRENBLUREL. 01, Last sequence update)

DT 01-MAR-2002 (TRENBLUREL. 20, Last annotation update)

DE Beta-lactamase (Fragment).

GN BLA.

OS Pseudomonas aeruginosa.

OC Plasmid PR01614.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI_TaxID=287;

RN (1)

RP SEQUENCE FROM N.A.

RA MEDLINE=95011664; PubMed=7926843;

RX West S.E., Schweizer H.P., Dell C., Sample A.K., Runyen-Tanecky L.J.;

RT *Construction of Improved Escherichia-Pseudomonas shuttle vectors

RT derived from pUC18/19 and sequence of the region required for their

RT replication in Pseudomonas aeruginosa.*;

RL Gene 148:81-86(1994).

EMBL; L30112; AAA66058.1; -

DR HSSP; P00810; 1XPB.

DR InterPro; IPR001466; Beta_lactamase.

DR InterPro; IPR000871; Beta_lactamase_A.

DR Pfam; PF00144; beta-lactamase; 1.

DR PRINTS; PR00118; BLACTAMASEA.

KW Plasmid.

FT NON_TER

SO SEQUENCE 105 AA; 11229 MW; D2889AA407330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;

Best Local Similarity 100.0%; Pred. No. 9; 1e-37;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 160 AAMTFLKLTGELTLASRQOLIDWNEADKVGPLRLSALPAGMTADKSGAGRGSRG 219

2 AAMTFLKLTGELTLASRQOLIDWNEADKVGPLRLSALPAGMTADKSGAGRGSRG 61

OY 220 IIAALGPDGKPSRIIVYITGSGATMDERNROIAGTASLIRKM 263
 |||||||
 Db 62 IIAALGPDGKPSRIIVYITGSGATMDERNROIAGTASLIRKM 105

RESULT 2

052330 PRELIMINARY; PRT: 105 AA.

AC 052330;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Tnl bla protea (Fragment).
 OS Escherichia coli.
 OC Plasmid RK2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=90264294; PubMed=2160936.
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT "The *kil*-kor regulon of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT *korc*.";
 RL J. Bacteriol. 173:3040-3050(1990).
 DR EMBL; M32794; AAA26408.1; -.
 DR HSSP; P00810; 1XPB.
 DR Interpro; IPR001466; Beta_lactamase.
 DR Interpro; IPR000871; Beta_lactamase_A.
 DR Pfam; PF00144; beta-lactamase; 1.
 DR PRINTS; PRO0118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER

SC SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 9.1e-37;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRKLLTGELTLASROQLIDMEADKYAGPLLSALPAGWFIADKSGAGERSRG 219
 |||||||
 Db 2 AMATTLRKLLTGELTLASROQLIDMEADKYAGPLLSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIIVYITGSGATMDERNROIAGTASLIRKM 263
 |||||||
 Db 62 IIAALGPDGKPSRIIVYITGSGATMDERNROIAGTASLIRKM 105

RESULT 3

08RTD8 PRELIMINARY; PRT: 145 AA.

AC 08RTD8;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE SHY-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Palasubramaniam S.;
 RT "Characterization of SHY-5 extended-spectrum beta-lactamase from
 RT ceftriaxone-resistant Klebsiella pneumoniae.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF467105; AAL75506.1; -.
 DR Interpro; IPR001466; Beta_lactamase.
 DR Pfam; PF00144; beta-lactamase; 1.
 FT NON_TER

FT NON_TER 145 145

SC SEQUENCE 145 AA; 15574 MW; F88634D619484C82 CRC64;

Query Match 36.7%; Score 493; DB 2; Length 145;
 Best Local Similarity 68.1%; Pred. No. 6e-34;
 Matches 94; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

OY 100 AATMSNDTANLLTITGKPELTFAPLNMGDHYRLDRWPEPELNEAIPNDRDTMPV 159
 |||||||
 Db 7 AATYMSNSAANLLTATVGGPAGLTAFRLQIDNTRLDRWTELNEALPGDARDTTTPA 66
 OY 160 AMATTLRKLLTGELTLASROQLIDMEADKYAGPLLSALPAGWFIADKSGAGERSRG 219
 |||||||
 Db 67 SMAATLRKLLTSGRLSARSQQLQMWVDDRVAGPLIRSVLPAGWFIADKSGAGERSRG 126

OY 220 IIAALGPDGKPSRIIVY 237
 |||||||
 Db 127 IIAALGPDGKPSRIIVY 144

RESULT 4

052639 PRELIMINARY; PRT: 102 AA.

AC 052639;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Beta-lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC Plasmid PRO1600.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]

RP SEQUENCE FROM N.A.
 RC TRANSPOSON-Tn1;
 RX MEDLINE=94336757; PubMed=8058819;
 RA Jansons I., Touchie G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT "Deletion and transposon mutagenesis and sequence analysis of the *pOF*
 RT PRO1600 *oriT* region found in the broad-host-range plasmids of the *pOF*
 RT series.";
 RL Plasmid 31:265-274(1994).
 DR EMBL; U22691; AAA8312.1; -.
 DR HSSP; P00810; 1XPB.
 DR Interpro; IPR001466; Beta_lactamase.
 DR Interpro; IPR000871; Beta_lactamase_A.
 DR Pfam; PF00144; beta-lactamase; 1.
 DR PRINTS; PRO0118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER

SC SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 9.6e-34;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRKLLTGELTLASROQLIDMEADKYAGPLLSALPAGWFIADKSGAGERSRG 219
 |||||||
 Db 2 AMATTLRKLLTGELTLASROQLIDMEADKYAGPLLSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIIVYITGSGATMDERNROIAGTASLIRKM 263
 |||||||
 Db 62 IIAALGPDGKPSRIIVYITGSGATMDERNROIAGTASLIRKM 105

RESULT 5

0990F3 PRELIMINARY; PRT: 138 AA.

AC 0990F3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-6988, and E/99 5-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327049; AK07464.1; -.
DR EMBL; AF327051; AK07466.1; -.
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;
SQ

Query Match 34.1%; Score 458; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 5.1e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTGGKELTAFILHNMGDHTRLDRWEPELNEAIPNDERDTTPVAMATTLRLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNTVRLDRWETELNEALPGARDTTPASMAATLRLT 69
OY 171 GELLTASROOLDIMWEADKVAGPLRSALPGWFIADKSGAGERSGCIITAAIGPDGP 230
DB 70 SQRLSARSQRLQWVDVAGPLIRSVLPAGWFIADKTSKRGARGIVALLGPNNKA 129
OY 231 SRIIVY 237
DB 130 ERIIVY 136

RESULT 6

O9AMA1 PRELIMINARY; PRT; 138 AA.
AC O9AMA1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1B/25;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327050; AK07465.1; -.
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;
SQ

Query Match 34.1%; Score 458; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 5.1e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTGGKELTAFILHNMGDHTRLDRWEPELNEAIPNDERDTTPVAMATTLRLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNTVRLDRWETELNEALPGARDTTPASMAATLRLT 69
OY 171 GELLTASROOLDIMWEADKVAGPLRSALPGWFIADKSGAGERSGCIITAAIGPDGP 230
DB 70 SQRLSARSQRLQWVDVAGPLIRSVLPAGWFIADKTSKRGARGIVALLGPNNKA 129
OY 231 SRIIVY 237
DB 130 ERIIVY 136

RESULT 7

O9AMA2 PRELIMINARY; PRT; 139 AA.
AC O9AMA2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-918;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327048; AK07463.1; -.
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS; PR00118; BLACTAMASEA.
FT NON_TER 1
FT SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;
SQ

Query Match 34.1%; Score 458; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 5.1e-31;
Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTGGKELTAFILHNMGDHTRLDRWEPELNEAIPNDERDTTPVAMATTLRLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNTVRLDRWETELNEALPGARDTTPASMAATLRLT 69
OY 171 GELLTASROOLDIMWEADKVAGPLRSALPGWFIADKSGAGERSGCIITAAIGPDGP 230
DB 70 SQRLSARSQRLQWVDVAGPLIRSVLPAGWFIADKTSKRGARGIVALLGPNNKA 129
OY 231 SRIIVY 237
DB 130 ERIIVY 136

RESULT 8

O9AMA0 PRELIMINARY; PRT; 138 AA.
AC O9AMA0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-E/98 9-1:
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.*;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327052; AK07467.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Query Match 33.9%; Score 456; DB 2; Length 138;
Best Local Similarity 66.9%; Pred. No. 7.5e-31;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTIGGPKELTAFLHNGDHYTRLDREPELNEAIPNDERDTTPVAMATTLRLT 170
DB 10 NLLATVGGPAGTAFALRQIDNTYRLDRWETELNEALPGDARDTTPASMAATLRLT 69
OY 171 GELLTLASRQQLDMWADKVAAGPLRSALPAGWFIADKSGAGERSGRTIAALGPDGKP 230
DB 70 SQRLSARSQROLQMWVDVAVAGPLIRSVLPAGWFIADKTKGASRGARGIALLGPNMKA 129
OY 231 SRIVIVY 237
DB 130 ERIVVLY 136

RESULT 9

O9AM99 PRELIMINARY; PRT; 138 AA.
ID O9AM99;
AC O9AM99;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/99 4-1;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.*;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327053; AK07468.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 33.9%; Score 456; DB 2; Length 138;
Best Local Similarity 66.9%; Pred. No. 7.5e-31;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTIGGPKELTAFLHNGDHYTRLDREPELNEAIPNDERDTTPVAMATTLRLT 170
DB 10 NLLATVGGPAGTAFALRQIDNTYRLDRWETELNEALPGDARDTTPASMAATLRLT 69
OY 171 GELLTLASRQQLDMWADKVAAGPLRSALPAGWFIADKSGAGERSGRTIAALGPDGKP 230
DB 70 SQRLSARSQROLQMWVDVAVAGPLIRSVLPAGWFIADKTKGASRGARGIALLGPNMKA 129

OY 231 SRIVIVY 237
DB 130 ERIVVLY 136

RESULT 10

O9AM98 PRELIMINARY; PRT; 139 AA.
ID O9AM98;
AC O9AM98;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/99 3-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.*;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327054; AK07469.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15234 MW; 0361A792F5442A8 CRC64;

Query Match 33.9%; Score 456; DB 2; Length 139;
Best Local Similarity 66.9%; Pred. No. 7.5e-31;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTIGGPKELTAFLHNGDHYTRLDREPELNEAIPNDERDTTPVAMATTLRLT 170
DB 10 NLLATVGGPAGTAFALRQIDNTYRLDRWETELNEALPGDARDTTPASMAATLRLT 69
OY 171 GELLTLASRQQLDMWADKVAAGPLRSALPAGWFIADKSGAGERSGRTIAALGPDGKP 230
DB 70 SQRLSARSQROLQMWVDVAVAGPLIRSVLPAGWFIADKTKGASRGARGIALLGPNMKA 129
OY 231 SRIVIVY 237
DB 130 ERIVVLY 136

RESULT 11

O9AM93 PRELIMINARY; PRT; 139 AA.
ID O9AM93;
AC O9AM93;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-935;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.*;
Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327047; AK07462.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.

DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SO SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Query Match
Best Local Similarity 66.9%; Pred. No. 1.4e-30;
Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NULLTTGGPKELTAFILHNGDHYTRLDREPELNEALPNDERDTTTPVAMATTLRLTLT 170
DB 10 HLLATVGGPAGLTAFLRQIGDNTYRLDRMETELNEALPGDARDTTTPASMAATTLRLTLT 69

OY 171 GELLTLASRQQLDWMADYVAGPLLSALAGMTADKSGAGERSGRTIALLGPDGR 230
DB 70 SORLSARSRQQLQMWVDVAGPLLSRVSIPAGMTADRTGASKRGARGIVALGPNKKA 129

OY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 12
053553 PRELIMINARY; PRT; 67 AA.

ID 053553
AC 053553
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OG Plasmid pMAN-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL; S81098; AAB35839.2; -.
DR HSSP; P00810; IXPB.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 67
SO SEQUENCE 67 AA; 7287 MW; 85F8C85B03r0CB CRC64;

Query Match
Best Local Similarity 98.5%; Pred. No. 2.5e-21;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 55 VLSRIDAGOEOLGRIRHYSQNLVESPTEKHLTDGMTVRELCSAAITMSDNTAAILL 114
DB 1 VLSRYDAGOEOLGRIRHYSQNLVESPTEKHLTDGMTVRELCSAAITMSDNTAAILL 60

OY 115 TTIGGPK 121
DB 61 TTIGGPK 67

RESULT 13
08GDES PRELIMINARY; PRT; 95 AA.
ID 08GDES
AC 08GDES;

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Bello H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Ames S.G.B.;
RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
Chile."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY150585; AAN7730.1; -.
FT NON_TER 1
FT NON_TER 95
SO SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
Best Local Similarity 64.2%; Pred. No. 1.3e-19;
Matches 61; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

OY 147 AIPNDERDTTTPVAMATTLRLTLTGELTLASRQQLDWMADYVAGPLLSALPAGMT 206
DB 1 ALPGARDTTTPASMAATTLRLTLTSORLSARSRQQLQMWVDVAGPLLSVLPAGMT 60

OY 207 ADKSGAGERSGRTIALLGPDGRPSRIVITYTGS 241
DB 61 ADKTGASRGARGIVALGPNKKAERIVITYTGS 95

RESULT 14
09JN58 PRELIMINARY; PRT; 62 AA.

ID 09JN58
AC 09JN58
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OG Plasmid pMAN-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE=96081517; PubMed=8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL; S81099; AAB35840.1; -.
DR HSSP; P00810; IXPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 62
SO SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
Best Local Similarity 98.4%; Pred. No. 1.6e-19;
Matches 61; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 132 DHYTRLDREPELNEALPNDERDTTTPVAMATTLRLTLTGELTLASRQQLDWMADYV 191
DB 1 DHYTRLDREPELNEALPNDERDTTTPVAMATTLRLTLTGELTLASRQQLDWMADYV 60

OY 192 AG 193
II

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

| | |
|----------------|---|
| Title: | SEQ2_37S_157T |
| Perfect score: | 1345 |
| Sequence: | 1 HPETLVKVKDAEDQLGARVQ.....TMDERNRQIAEIGASLIKHW 263 |

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```

Searched:      1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 865366

```

```

Minimum DB seq length: 0
Maximum DB seq length: 207
Post-processing: Minimum M

```

Database :

1: A_Geneseq_19Jun03.*

1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980. DAT.*

2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981. DAT.*

3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982. DAT.*

4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983. DAT.*

5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984. DAT.*

6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985. DAT.*

7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986. DAT.*

8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987. DAT.*

9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988. DAT.*

10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989. DAT.*

11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990. DAT.*

12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991. DAT.*

13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992. DAT.*

14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993. DAT.*

15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994. DAT.*

16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995. DAT.*

17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996. DAT.*

18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997. DAT.*

19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998. DAT.*

20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999. DAT.*

21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000. DAT.*

22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001. DAT.*

23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002. DAT.*

24: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2003. DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 583 | 43.3 | 196 | 7 | AAP60627 | Beta-urogastrone - |
| 2 | 567 | 42.2 | 146 | 22 | AAU23221 | Novel human enzyme |
| 3 | 567 | 42.2 | 146 | 22 | AAM90870 | Human immune/haema |
| 4 | 488 | 36.3 | 127 | 18 | AAW20440 | H. pylori cytolass |
| 5 | 473 | 35.2 | 129 | 22 | AAU22281 | Novel human enzyme |
| 6 | 473 | 35.2 | 129 | 22 | AAM90871 | Human immune/haema |
| 7 | 473 | 35.2 | 182 | 22 | ABG27917 | Novel human diagno |
| 8 | 457 | 34.0 | 94 | 21 | AAB59052 | Breast and ovarian |
| 9 | 400 | 29.7 | 159 | 7 | AAP60628 | Beta-urogastrone - |

| | | | | | | |
|----|------|------|-----|----|-----------|----------------------|
| 10 | 366 | 27.2 | 101 | 22 | ABG37935 | Novel human diageno |
| 11 | 329 | 24.5 | 88 | 22 | ABG27919 | Novel human diageno |
| 12 | 195 | 14.5 | 119 | 7 | AAp60626 | Beta-urogastrone - |
| 13 | 145 | 10.8 | 51 | 21 | AAAY92783 | Viggs-beta-lactama |
| 14 | 83 | 6.2 | 200 | 18 | AAW55524 | H. pylori ORF 296p |
| 15 | 80.5 | 6.0 | 158 | 22 | AAU45158 | Proptonibacterium |
| 16 | 75 | 5.6 | 202 | 24 | AAE32763 | Human zcyto24 prot |
| 17 | 75 | 5.6 | 202 | 24 | AAO16283 | Mouse IMX129840-2 |
| 18 | 72.5 | 5.4 | 196 | 21 | AAAY3344 | HRM clone 0258181 |
| 19 | 71 | 5.3 | 159 | 19 | AAW69170 | N-terminally tagged |
| 20 | 70.5 | 5.2 | 193 | 23 | AAU76854 | Human integrin alp |
| 21 | 70.5 | 5.2 | 193 | 23 | AAU76853 | Human integrin alp |
| 22 | 70 | 5.2 | 15 | 10 | AAAP98503 | Sequence encoded b |
| 23 | 70 | 5.2 | 170 | 22 | AAU52472 | Proptonibacterium |
| 24 | 69.5 | 5.2 | 170 | 22 | AAU56330 | Proptonibacterium |
| 25 | 69.5 | 5.2 | 184 | 21 | AAE25446 | Plnus radiata cell |
| 26 | 69 | 5.1 | 134 | 21 | AAE37220 | Plus mayis protein f |
| 27 | 69 | 5.1 | 166 | 23 | ABP26537 | Streptococcus poly |
| 28 | 68 | 5.1 | 194 | 22 | AAAG81764 | S. epidermidis ope |
| 29 | 68 | 5.1 | 180 | 21 | AAE20205 | Arabidopsis thaila |
| 30 | 67.5 | 5.0 | 152 | 22 | ABG19118 | Novel human diageno |
| 31 | 67.5 | 5.0 | 152 | 22 | ABG27771 | Novel human diageno |
| 32 | 67.5 | 5.0 | 187 | 23 | AAO97104 | Human MK61 protein |
| 33 | 67 | 5.0 | 178 | 22 | AAAG91926 | C glutamicum prote |
| 34 | 67 | 5.0 | 202 | 22 | AAE82686 | S. epidermidis ope |
| 35 | 66.5 | 4.9 | 83 | 22 | AAU59092 | Proptonibacterium |
| 36 | 66.5 | 4.9 | 177 | 21 | AAE41582 | S. pneumoniae deri |
| 37 | 66 | 4.9 | 183 | 19 | AAAY85905 | Human ORFX ORF1346 |
| 38 | 66 | 4.9 | 202 | 24 | AAE32764 | S. pneumoniae deri |
| 39 | 65.5 | 4.9 | 50 | 17 | AAU03551 | Human zcyto25 prot |
| 40 | 65.5 | 4.9 | 136 | 21 | AAU44229 | Human alpha 2 C4 a |
| 41 | 65 | 4.8 | 150 | 21 | AAAY95249 | Partial corn extra |
| 42 | 65 | 4.8 | 168 | 21 | AAAG58698 | Soybean calmodulin |
| 43 | 65 | 4.8 | 203 | 18 | AAW31365 | Arabidopsis thaila |
| 44 | 65 | 4.8 | 203 | 22 | ABO60157 | Neospora caninum a |
| 45 | 65 | 4.8 | 207 | 22 | ABG19320 | Novel human diageno |

ALIGNMENTS

| | |
|----------|--|
| RESULT | 1 |
| AAP60627 | |
| ID | AAP60627 standard; Protein; 196 AA. |
| XX | |
| AC | AAP60627; |
| XX | |
| DT | 25-MAR-2003 (updated) |
| DT | 17-JUN-1991 (first entry) |
| XX | |
| XX | Beta-urogastrone - beta-lactamase fusion protein from pUG2101. |
| KM | Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds |
| KW | fusion protein; beta-lactamase. |
| XX | |
| OS | Synthetic. |
| XX | |
| FH | Key |
| FT | Protein |
| FT | 1..120 location/Qualifiers |
| FT | /label= beta-lactamase |
| FT | Misc-difference 121..123 |
| FT | /label= adaptor |
| FT | 124..196 |
| FT | /label= beta-urogastrone |
| PN | DE3523634-A. |
| XX | |
| PD | 09-JAN-1986. |
| XX | |
| PE | 02-JUL-1985; 85DE-3523634. |
| PK | 02-JUL-1984; 84JP-0137691. |
| XX | |

PA (EART) EARTH CHEM CO LTD.
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
PI Matsushiro S;
XX WPI: 1986-015031/03.
DR N-PSDB: AAN60631.
XX
PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PT and transformed cells contg. it.
XX
PS Disclosure: Page 56-59; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
CC
XX
SQ Sequence 196 AA;
Query Match 43.3%; Score 583; DB 7; Length 196;
Best Local Similarity 98.3%; Pred. No. 7.6e-52;
Matches 115; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 HPEPTLVKVNDAEQLGARVGIETDLSNGKILSPRSERPEPMSTFKYLLCGAVSRID 60
DB 24 HPEPTLVKVNDAEQLGARVGIETDLSNGKILSPRSERPEPMSTFKYLLCGAVSRVD 83
QY 61 AGGEQLGRRIHYSONDLVEVSPYTEKHILTDGMTVRELCSAATITMSDNFANILLTTI 117
DB 84 AGGEQLGRRIHYSONDLVEVSPYTEKHILTDGMTVRELCSAATITMSDNFANILLTTI 140
RESULT 2
AAU23221
ID AAU23221 standard; Protein: 146 AA.
XX
AC AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotoxic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
PD
XX
PD 02-AUG-2001.
PF
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227709.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMAN-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
DR WPI; 2001-465566/50.
DR N-PSDB; AAS41091.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases
XX
XX
PS Claim 11; SEQ ID NO 1217; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the

CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC A022915-AA023814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 146 AA:

Query Match 42.2%; Score 567; DB 22; Length 146;
Best Local Similarity 97.4%; Pred. No. 2,2e-50;
Matches 112; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 HPETLVKVKADQKARKVGYIELDLSGKILSFSESESPPMWTFKVLGAYLRID 60
|||
Db 28 HPETLVKVKADQKARKVGYIELDLSGKILSFSESESPPMWTFKVLGAYLRID 87

QY 61 AGQEQLGRRIRHYSQNDLYEYSPYTERHLDGKMTVRELCSAATWSDNTAAILLT 115
|||
Db 88 AGQEQLGRRIRHYSQNDLYEYSPYTERHLDGKMTVRELCSAATWSDNTAAILLT 142

RESULT 3
ID AAM90870 standard; Protein; 146 AA.
AC AAM90870;
XX
XX
DT 07-NOV-2001 (first entry)
XX
XX
DE Human immune/haematopoietic antigen SEQ ID NO:18463.
XX
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytosolic; gene therapy; vaccine; metastasis.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200157182-A2.
PD
PD 09-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0235802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.
PR 08-NOV-2000; 2000US-0244677.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249307.
PR 17-NOV-2000; 2000US-0249308.
PR 17-NOV-2000; 2000US-0249309.
PR 17-NOV-2000; 2000US-0249310.
PR 17-NOV-2000; 2000US-0249311.
PR 17-NOV-2000; 2000US-0249312.
PR 17-NOV-2000; 2000US-0249313.
PR 17-NOV-2000; 2000US-0249314.
PR 17-NOV-2000; 2000US-0249315.
PR 17-NOV-2000; 2000US-0249316.
PR 17-NOV-2000; 2000US-0249317.
PR 17-NOV-2000; 2000US-0249318.
PR 17-NOV-2000; 2000US-0249319.
PR 17-NOV-2000; 2000US-0249320.
PR 17-NOV-2000; 2000US-0249321.
PR 17-NOV-2000; 2000US-0249322.
PR 17-NOV-2000; 2000US-0249323.
PR 17-NOV-2000; 2000US-0249324.
PR 17-NOV-2000; 2000US-0249325.
PR 17-NOV-2000; 2000US-0249326.
PR 17-NOV-2000; 2000US-0249327.
PR 17-NOV-2000; 2000US-0249328.
PR 17-NOV-2000; 2000US-0249329.
PR 17-NOV-2000; 2000US-0249330.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251857.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251899.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-465566/50.
N-PSDB: AAS41090.

PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases

PS Claim 11; SEQ ID NO 1216; 1180bp; English.

XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC .AAS2915-AAS29814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 129 AA:

Query Match 35.2%; Score 473; DB 22; Length 129;

Best Local Similarity 96.9%; Pred. No. 9.7e-41;
Matches 93; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 HPELTVKVKDAEDQAGARVGYIELDLSNGKILSFSESEEPKPMSTFKYLCAVLSTRID 60
Db 28 HPELTVKVKDAEDQAGARVGYIELDLSNGKILSFSESEEPKPMSTFKYLCAVLSTRID 87
QY 61 AGOEQLGRIRIHSNDLVESPYTEKHLTDGMVIRE 96
Db 88 AGOEQLGRIRIHSNDLVESPYTEKHLTDGMVIRE 123

RESULT 6
ID AAM90871 standard; Protein; 129 AA.
XX AAM90871;
AC AAM90871;
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/Haematopoietic antigen SEQ ID NO:18464.
XX
KW Human; Immune; haematopoietic; Immune/Haematopoietic antigen; cancer;
XX cytosolic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX
PN M0200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 02-MAR-2000; 2000US-0184664.
PR 16-MAR-2000; 2000US-0186350.
PR 17-MAR-2000; 2000US-0189874.
PR 18-APR-2000; 2000US-0190076.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239393.
PR 13-OCT-2000; 2000US-0239397.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.


```

XX AAB59052;
AC
XX 27-MAR-2001 (first entry)
XX
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antifungal; antifungal; antiviral; anticonvulsant;
XX antidiabetic; antifungal; antiparasitic; cardiac; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
XX Homo sapiens.
XX
XX NC020055173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05881.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-611515/58.
XX
XX N-PSDB; AAF21955.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Claim 11; Page 1228; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antifungal; antiviral; vulnery; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiac activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemias; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX
XX Sequence 94 AA;
SQ
Query Match 34.0%; Score 457; DB 21; Length 94;
Best Local Similarity 97.8%; Pred. No. 2.8e-35;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

RESULT 9
AAP60628
ID AAP60628 standard; Protein; 159 AA.
XX
XX AAP60628;
AC
XX 25-MAR-2003 (updated)
XX 17-JUN-1991 (first entry)
XX
DE Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX
XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX fusion protein; beta-lactamase.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Protein 1..96
XX Misc-difference /label= beta-lactamase
XX /label= adaptor
XX Protein 101..159
XX /label= beta-urogastrone
XX
XX DE3523634-A.
XX
XX 09-JAN-1986.
XX
XX 02-JUL-1985; 85DE-3523634.
XX
XX 02-JUL-1984; 84JP-0137691.
XX
XX (EART) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgai H, Horioka A, Hiramatsu H, Koumoto S, Mishimura A;
XX Matsushiro S;
XX WPI: 1986-015031/03.
XX
XX N-PSDB; AAN60632.
XX
XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
XX and transformed cells contg. it.
XX
XX Disclosure; Page 59-61; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
XX protects beta-urogastrone and beta-lactamase collects in the periplasm
XX of E. coli. It is therefore easy to collect and purify the product.
XX Beta-urogastrone is the hormone of the salivary glands which suppresses
XX stomach acid secretion and promotes cell growth, so is useful for
XX treating ulcers and wounds. Previously the product was obt. only
XX in small amts. from human urine.
XX See also AAN60628, and 30-32.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 159 AA;
SQ
Query Match 29.7%; Score 400; DB 7; Length 159;
Best Local Similarity 68.3%; Pred. No. 4.6e-33;
Matches 86; Conservative 8; Mismatches 20; Indels 12; Gaps 2;

```

```

DB      140 VVCYIG 145

RESULT 10
ABG27935
ID      ABG27935 standard; Protein: 101 AA.
XX      AC
XX      ABG27935;
XX      DT
XX      18-FEB-2002 (first entry)
XX      DE
XX      Novel human diagnostic protein #27926.
XX      KW
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX      food supplement; medical imaging; diagnostic; genetic disorder.
XX      OS
XX      Homo sapiens.
XX      PN
XX      W0200175067-A2.
XX      PD
XX      11-OCT-2001.
XX      PF
XX      30-MAR-2001; 2001WO-US08631.
XX      PR
XX      31-MAR-2000; 2000US-0540217.
XX      PR
XX      23-AUG-2000; 2000US-0649167.
XX      PA
XX      (HYSE-) HYSEQ INC.
XX      PI
XX      Dmanac RT, Liu C, Tang YT;
XX      DR
XX      WPI; 2001-639362/73.
XX      DR
XX      N-PSDB; MAS92122.

PT      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity -
PS      Claim 20; SEQ ID No 58294; 103bp; English.

XX      CC
XX      The invention relates to isolated polynucleotide (I) and
XX      CC
XX      polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX      CC
XX      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX      CC
XX      and gene mapping, and in recombinant production of (II). The
XX      CC
XX      polynucleotides are also used in diagnostics as expressed sequence tags
XX      CC
XX      for identifying expressed genes. (I) is useful in gene therapy techniques
XX      CC
XX      to restore normal activity of (II) or to treat disease states involving
XX      CC
XX      (II). (II) is useful for generating antibodies against it, detecting or
XX      CC
XX      quantitating a polypeptide in tissue, as molecular weight markers and as
XX      CC
XX      a food supplement. (II) and its binding partners are useful in medical
XX      CC
XX      imaging of sites expressing (II). (I) and (II) are useful for treating
XX      CC
XX      disorders involving aberrant protein expression or biological activity.
XX      CC
XX      The polypeptide and polynucleotide sequences have applications in
XX      CC
XX      diagnostics, forensics, gene mapping, identification of mutations
XX      CC
XX      responsible for genetic disorders or other traits to assess biodiversity
XX      CC
XX      and to produce other types of data and products dependent on DNA and
XX      CC
XX      amino acid sequences. ABG00010-ABG30377 represent novel human
XX      CC
XX      diagnostic amino acid sequences of the invention.
XX      CC
XX      Note: The sequence data for this patent did not appear in the printed
XX      CC
XX      specification, but was obtained in electronic format directly from WIPO
XX      CC
XX      at ftp.wipo.int/pub/published_sequences.
XX      SQ
XX      Sequence      101 AA;

Query Match      27.2%; Score 366; DB 22; Length 101;
Best Local Similarity 98.6%; Pred. No. 7, 8e-30;
Matches 73; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

154 DTTTPAAMATTKRLTGTGELLTTLASRQOLIDMEADKVGGLRLSALPAGFIIDKSGAG 213
|||||
10 DTTTPAAMATTKRLTGTGELLTTLASRQOLIDMEADKVGGLRLSALPAGFIIDKSGAG 69

```

0Y 214 ERGSRGITIALGPD 227
|||||
Db 70 ERGSRGITIALGPD 83

RESULT 11
ABG27919
ID ABG27919 standard; Protein; 88 AA.
AC ABG27919;
XX
XX 18-FEB-2002 (first entry)
DT
XX Novel human diagnostic protein #27910.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
PN WO200175067-A2.
PP
XX 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSED INC.
XX
PI Drmanac RT, Liu C, Yang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSTB; AAS92106.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20: SEQ ID NO 58278; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABC00010-ABC30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 88 AA;

Query Match 24.5%; Score 329; DB 22; Length 88;
Best Local Similarity 81.2%; Pred. No. 4; 3e-26;
Matches 69; Conservative 2; Mismatches 4; Indels 10; Gaps 1

4 TLVKKDAEDQLGARVGITIEDLMSGKLTLESRRSEERFPWMTTFKVLICAVISRDAQ 63
|||||

Db 1 TLVVKDAEDDQAGARVGYIELDNSGKILSFREPERPMSTFKVLLCGAVLSRIDAG 60
 QY 64 EOLGRRI-----HSQNDLV 78
 Db 61 EOLGRRIILRLMTWLSTHOSKSL 85

RESULT 12

AAP60626
 ID AAP60626 standard; Protein; 119 AA.

XX AAP60626;

XX 25-MAR-2003 (updated)

DT 17-JUN-1991 (first entry)

DE Beta-urogastrone - beta-lactamase fusion protein from pUC2301.

XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;

KW fusion protein; beta-lactamase.

XX Synthetic.

XX Key Location/Qualifiers

FT Protein 1..63

FT Misc-difference 64..66 /label= beta-lactamase

FT Protein 67..119 /label= adaptor

FT Protein /label= beta-urogastrone

XX DE3523634-A.

XX 09-JAN-1986.

XX 02-JUL-1985; 85DE-3523634.

XX 02-JUL-1984; 84JP-0137691.

XX (EMRT) EARTH CHEM CO LTD.

XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;

XX Matsushiro S;

XX WPI: 1986-015031/03.

XX N-PSDB; AAN60630.

XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)

PT and transformed cells contg. it.

XX Disclosure; Page 55; 92pp; German.

XX The fusion protein is less easily degraded by proteases and so

CC protects beta-urogastrone and beta-lactamase collects in the periplasm

CC of E. coli. It is therefore easy to collect and purify the product.

CC Beta-urogastrone is the hormone of the salivary glands which suppresses

CC stomach acid secretion and promotes cell growth, so is useful for

CC treating ulcers and wounds. Previously the product was obtd. only

CC in small amts. from human urine.

CC See also AAN60628, and 30-32.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 119 AA;

Query Match 14.5%; Score 195; DB 7; Length 119;

Best Local Similarity 97.5%; Pred. No. 4.7e-12; Mismatches 1; Indels 0; Gaps 0;

Db 1 HPTLVKVKDAEDDQAGARVGYIELDNSGKILSFREPER 40
 24 HPTLVKVKDAEDDQAGARVGYIELDNSGKILSFREPER 63

RESULT 13
 AAY92783
 ID AAY92783 standard; Protein; 51 AA.

XX AAY92783;

XX 29-AUG-2000 (first entry)

DT Vtgs-beta-lactamase fusion protein (partial).

XX Vtgs; vitellogenin; secretory signal sequence; gene expression;

KW oestrogen receptor binding protein; systemic circulation;

XX beta-lactamase.

XX Chimeric - Oreochromis aureus.

XX Chimeric - Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..21

FT Cleavage-site 15..16 /label= Vtgs

FT Protein 22..51

XX /note= "beta-lactamase mature protein"

XX WO200026366-A1.

XX 11-MAY-2000.

XX 29-OCT-1999; 99WO-SG00108.

XX 30-OCT-1998; 98US-0106426.

XX 26-OCT-1999; 99US-0426776.

XX (VYSI-) UNIV SINGAPORE NAT.

XX (LAMT/) LAM T J.

XX Ding JL, Tan NS, Ho B;

XX WPI: 2000-365615/31.

XX N-PSDB; AAA28507.

XX Isolated nucleic acid for assaying for heterologous gene expression,

PT detecting presence of compound that binds to estrogen receptor or

PT encoding secretory signal sequence

XX Example 6; Fig 14A; 73pp; English.

XX A reporter beta-lactamase system that uses the Oreochromis aureus,

CC vitellogenin secretory sequence (Vtgs), designated pBAYtghlactkana was

CC constructed. Vtgs and variants that comprise conservative

CC replacements that retain the biological activities of directing secretion

CC of a fusion protein from a cell and cleavage of the secretory signal

CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be

CC fused to either a reporter protein or a lipopolysaccharide-binding

CC protein coding sequence. The isolated nucleic acid is useful for assaying

CC for heterologous gene expression, detecting the presence of a compound

CC that binds to an oestrogen receptor in a sample or producing a desired

CC protein from a host cell. It can also be used in a method for obtaining

CC systemic circulation of a desired protein in a transgenic or chimeric

CC host organism.

XX Sequence 51 AA;

Query Match 10.8%; Score 145; DB 21; Length 51;

Best Local Similarity 100.0%; Pred. No. 2e-07; Mismatches 0; Indels 0; Gaps 0;

Db 2 PETLVKVKDAEDDQAGARVGYIELDNSGK 30
 23 PETLVKVKDAEDDQAGARVGYIELDNSGK 51

RESULT 14
ID AAW55524 standard; Protein: 200 AA.
XX AC AAW55524;
XX DT 02-JUL-1998 (first entry)
XX DE H. pylori ORF 29epI0720_24432762_c3_39 cellular protein.
XX KW Cytoplasmic; vaccine; prevention; infection; treatment; envelope;
XX RW Identification; binding compound; bacteria; life cycle; activator;
XX OS Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX Helicobacter pylori.
XX PR W09737044-A1.
XX PD 09-OCT-1997.
XX PE 27-MAR-1997; 97WO-US05223.
XX PR 06-DEC-1996; 96US-0761318.
XX PR 29-MAR-1996; 96US-0625811.
XX PR 02-APR-1996; 96US-0758731.
XX PR 25-OCT-1996; 96US-0736905.
XX PR 28-OCT-1996; 96US-0738859.
XX PA (ASTR) ASTRA AB.
XX PI Alm RA, Smith D;
XX WPI: 1997-503122/46.
XX DR N-PSDB: AAV24933.
XX PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX PS Claims 14,93; Page 726-727; 1145pp; English.
XX This sequence is a H. pylori cellular protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX SQ Sequence 200 AA;

Query Match 6.2%; Score 83; DB 18; Length 200;
Best Local Similarity 25.9%; Pred No. 3.7;
Matches 51; Conservative 31; Mismatches 43; Indels 72; Gaps 13

22 IEIDLNSGKILIESE--RSEERFPMMSTFKVLLCGAVLSRIDAGQEQIGRRIRHYSQNDLV 78
:||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db VLDITDQSOKSMETATATAKEKPFFSLF-----NSSGSDDLQAKWVKENILI 61
79 -----EYSPVTEKHLLTDCMTVRELCSAAITWSDMTAANLLL--TTIGGPKEI-TAFLEHNMG 131

DB 62 DKEGYSKETOK-----AMLS-----NIVLPPT--PSQLOTEVIANML 99
OY 132 DHVTRIDRWEPELEINAIIPNDEROTTPVAMATLRKILT-----CELLFLASRQOLIDME 187
DB 100 EREEOQ-----ELNE-----NIRALIVINRMPRIIPPLKEKQALIEFIK 138
OY 188 ---ADKVAGPILRSAL 200
DB 139 ENNPSPDRIT--LLESSL 153
RESULT 15
AAU45138
ID AAU45138 standard; Protein; 158 AA.
XX AAU45138;
XX
XX
XX 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #6034.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001MO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'walsonneuve J, Zhang Y, Jen S, Carter D;
XX
XX MPI: 2001-616774/71.
XX DR N-PSDB; AAS59525.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX
XX Example 1; SEQ ID No 6333; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 Seconds

(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_37S_157T

Perfect score: 1345
Sequence: 1 HPETLVKVKDAEDQLGARVY.....TMDERNRQIAIEIGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5
Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|-------------------|
| 1 | 74.5 | 5.5 | 124 | US-08-311-731A-202 | Sequence 202, App |
| 2 | 72.5 | 5.4 | 197 | US-09-252-991A-24241 | Sequence 24241, A |
| 3 | 71 | 5.3 | 159 | US-08-991-890-4 | Sequence 4, Appl1 |
| 4 | 69 | 5.1 | 169 | US-09-328-352-4491 | Sequence 4491, Ap |
| 5 | 65.5 | 4.9 | 158 | US-09-010-809-19 | Sequence 19, Appl |
| 6 | 65 | 4.8 | 150 | US-09-239-909-2 | Sequence 2, Appl1 |
| 7 | 64.5 | 4.8 | 178 | US-09-252-991A-29942 | Sequence 29942, A |
| 8 | 64.5 | 4.8 | 198 | US-09-413-814-87 | Sequence 87, Appl |
| 9 | 64 | 4.8 | 203 | US-08-624-677A-2 | Sequence 2, Appl1 |
| 10 | 63.5 | 4.7 | 144 | US-08-225-480-4 | Sequence 4, Appl1 |
| 11 | 63.5 | 4.7 | 144 | US-09-118-445-4 | Sequence 4, Appl1 |
| 12 | 63.5 | 4.7 | 160 | US-09-252-991A-24737 | Sequence 24737, A |
| 13 | 63.5 | 4.7 | 170 | US-08-858-207A-519 | Sequence 519, App |
| 14 | 62 | 4.6 | 132 | US-09-252-991A-22681 | Sequence 22681, A |
| 15 | 62 | 4.6 | 181 | US-08-482-142-105 | Sequence 105, App |
| 16 | 62 | 4.6 | 181 | US-08-478-572-195 | Sequence 195, App |
| 17 | 62 | 4.6 | 181 | US-08-484-296-195 | Sequence 195, App |
| 18 | 61.5 | 4.6 | 142 | US-09-345-473E-31 | Sequence 31, Appl |
| 19 | 61 | 4.5 | 144 | US-08-557-122A-6 | Sequence 25578, A |
| 20 | 61 | 4.5 | 174 | US-08-557-122A-6 | Sequence 6, Appl1 |
| 21 | 61 | 4.5 | 174 | US-08-557-122A-6 | Sequence 6, Appl1 |
| 22 | 61 | 4.5 | 189 | US-08-671-548C-48 | Sequence 48, Appl |
| 23 | 61 | 4.5 | 200 | US-08-557-122A-12 | Sequence 12, Appl |
| 24 | 61 | 4.5 | 200 | US-08-557-122A-12 | Sequence 12, Appl |
| 25 | 60.5 | 4.5 | 103 | US-09-262-666-12 | Sequence 1262, Ap |
| 26 | 60.5 | 4.5 | 167 | US-08-690-849-2 | Sequence 2, Appl1 |
| 27 | 60.5 | 4.5 | 167 | US-09-004-053-2 | Sequence 2, Appl1 |

| | | | | | |
|----|------|-----|-----|----------------------|-------------------|
| 28 | 60 | 4.5 | 145 | US-09-134-001C-4982 | Sequence 4982, Ap |
| 29 | 59.5 | 4.4 | 146 | US-09-134-001C-5269 | Sequence 5269, Ap |
| 30 | 59 | 4.4 | 171 | US-09-107-532A-3979 | Sequence 3979, Ap |
| 31 | 59 | 4.4 | 180 | US-09-194-905-5 | Sequence 5, Appl1 |
| 32 | 59 | 4.4 | 199 | US-09-252-991A-30363 | Sequence 30363, A |
| 33 | 58.5 | 4.3 | 203 | US-09-252-991A-24921 | Sequence 24921, A |
| 34 | 58 | 4.3 | 136 | US-09-252-991A-30622 | Sequence 30622, A |
| 35 | 58 | 4.3 | 168 | US-08-451-947-6 | Sequence 6, Appl1 |
| 36 | 58 | 4.3 | 168 | US-08-424-826A-6 | Sequence 6, Appl1 |
| 37 | 58 | 4.3 | 168 | US-08-928-694-6 | Sequence 6, Appl1 |
| 38 | 58 | 4.3 | 168 | US-08-450-842-6 | Sequence 6, Appl1 |
| 39 | 58 | 4.3 | 168 | US-08-451-390-6 | Sequence 6, Appl1 |
| 40 | 58 | 4.3 | 168 | PCT-US91-06950-6 | Sequence 6, Appl1 |
| 41 | 58 | 4.3 | 183 | US-09-252-991A-31347 | Sequence 31347, A |
| 42 | 58 | 4.3 | 187 | US-09-252-991A-29325 | Sequence 29325, A |
| 43 | 58 | 4.3 | 189 | US-09-252-991A-17160 | Sequence 17160, A |
| 44 | 58 | 4.3 | 204 | US-09-252-991A-17611 | Sequence 17611, A |
| 45 | 58 | 4.3 | 206 | US-09-029-603-3 | Sequence 3, Appl1 |

ALIGNMENTS

RESULT 1

US-08-311-731A-202
Sequence 202, Application US/08311731A
Patent No. 6583266

GENERAL INFORMATION:

APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESSES:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311, 731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:

SEQUENCE CHARACTERISTICS:

LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae

US-08-311-731A-202

Query Match 5.5%; Score 74.5; DB 4; Length 124;

Best Local Similarity 28.6%; Pred. No. 1.8;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;

QY 130 MGDHV--TRUDREPELN---EAIIPNDERD-----TFVVAATIRKLITGLTGLTLASQ 180

DB 12 MDDGIGMEREGMTGTCOPLRVVPGDESTLDCRASPELIT--NLSPTIMHPPPS 69
QY 181 QUIDMMEA-DKVAAGPL-----RSALPAGWFIADKSGAGRGSGIIALOPDGKPSR 232
DB 70 RDDDVPEPDALOGTAVFPAATGKATPAAVGI-----GASTRGS-GILASLSPFQPAR 123

RESULT 2
US-09-252-991A-24241
; Sequence 24241, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24241
; LENGTH: 197
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match 5.4%; Score 72.5; DB 4; Length 197;
Best Local Similarity 23.6%; Pred. No. 6.1;
Matches 37; Conservative 23; Mismatches 42; Indels 55; Gaps 7;

QY 94 VHELCSAATMSDNTANLLTTIGCPKELTAPLHMGDHYTRLDRWPELNEAIPNDER 153
DB 12 VMLSSARTSASLASVPLPIATP---TSARPSASLT-----PSPVI 55
QY 154 DTTTPVA-MATTLRLKLTGLTTLA-----SRQULDMWADKVAAP----- 194
DB 56 ATTSPLACRACRTRSLCSG--LARANSTPOSTSR--WSSNSISAPVSGRSMPTP 110
QY 195 -----LRSALPAGWFIADKSGAGERG 216
DB 111 SMAPIAAVSTWSPVILFTALPAWHSMTATATASSRG 147

RESULT 3
US-08-991-890-4
; Sequence 4, Application US/08991890
; Patent No. 6114307
; GENERAL INFORMATION:
; APPLICANT: Jaepers, Stephen R.
; APPLICANT: Sprugel, Katherine H.
; APPLICANT: Ren, Hong Ping
; APPLICANT: Humes, Jacqueline M.
; APPLICANT: Hoffman, Ross C.
; APPLICANT: Conklin, Bartell C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: STIMULATING PANCREATIC ISLET CELL REGENERATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,890
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/033,003
; FILING DATE: December 16, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 96-41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-991-890-4

Query Match 5.3%; Score 71; DB 3; Length 159;
Best Local Similarity 23.8%; Pred. No. 6.5;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

QY 49 VLCCGAVLSRIDNGE-----QUGRIHYSON-----DLVEYSPTTEKH 87
DB 12 LLLCGAVP--VSPSOEIAEFORGRHHHHHGGSGAELGCGFRGKHLSTCPMPKPT 69
QY 88 LIDGWTVELCSAATMSDNTANLLTTIGCPKELTAPLHMGDHYTRLDRWPELNEA 147
DB 70 FT-----TTGCGWLES--GRKEMVSTSNKRGQAL-----GTSEF 105
QY 148 IPNDERDTPVAMA-TTLRLKL 169
DB 106 IPNLSPELKKPLSEGPSLKKII 128

RESULT 4
US-09-328-352-4491
; Sequence 4491, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4491
; LENGTH: 169
; TYPE: PRF
; ORGANISM: Acinetobacter baumannii
US-09-328-352-4491

Query Match 5.1%; Score 69; DB 4; Length 169;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 36; Conservative 34; Mismatches 60; Indels 32; Gaps 7;

QY 118 GGPKELTAPLHMGDHYTRLDRWPELNEAIPNDERDTT-----TPVAMATTLRLKL 169
DB 5 GGEETIMKMLAQON---RFDQWEQIVQYL---DLTVREIRINVVFTIIVVVIVGYS 58
QY 170 TGEHLTLASRQO-----LIDWMEADKVAAGPLRLSALPAGWFIADKSGAGERG--GSRG 219
DB 59 LMKMSLAEQQOQRLNDKDLMMVMOQNAV-----TKKPANELDRSGKIORVAQDQG 112
QY 220 IIAAGPDGKPSRIIVYITGSGQATMDERNQIAEIGASLAK 261

Db 113 LTVSSQONGEQLQIVV--THONTAIIANLTQIAOMGLSTICK 152

RESULT 5

US-09-010-809-19
; Sequence 19, Application US/09010809B
; Patent No. 6090601
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Beljach, Mary C.
; TITLE OF INVENTION: Epithelione Polyketide Synthases and Encoding DNA
; TITLE OF INVENTION: Theoretor
; FILE REFERENCE: 30062-20020.00
; CURRENT APPLICATION NUMBER: US/09/010,809B
; CURRENT FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-010-809-19

Query Match 4.9%; Score 65.5; DB 3; Length 158;
Best Local Similarity 27.6%; Pred. No. 26;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

QY 106 DNTAANLLT-----TIGPKELTAFLHNNGDHT--RLDRWED----ELNEAIPNDERD 154
Db 21 NHDARKHLLTSROGASAPGADVLRSELALGASVTIACDVADPRALKDLIDNIP----- 75

QY 155 TTTVAAMATTLTKLTGELLTLASRQQLIDMMEADKYAGLRSNA-----LPAQWF 205
Db 76 SAHVVAANVRAASVADGDLGAMSLERT-----DRVFAKIDAAWHLHOTQDKPLAAF 129

QY 206 IADSGAGERSRG 219
Db 130 ILFSSVAGVAGSSG 143

RESULT 6
US-09-239-909-2
; Sequence 2, Application US/09239909
; Patent No. 6284952
; GENERAL INFORMATION:
; APPLICANT: Kumho Petrochemical Co. Ltd.
; TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve M
; FILE REFERENCE: P99-2-6
; CURRENT APPLICATION NUMBER: US/09/239,909
; EARLIER APPLICATION NUMBER: 1999-01-29
; EARLIER FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: KOPATIN 1.0
; SEQ ID NO 2
; LENGTH: 150
; TYPE: PRT
; ORGANISM: G. max calmodulin4 (scam4)
US-09-239-909-2

Query Match 4.8%; Score 65; DB 3; Length 150;
Best Local Similarity 27.9%; Pred. No. 27;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

QY 73 SONDLVESPVTETKHLTNG---MTVRELSAATMSDNTAANLLTTI-----GGPKF 122
Db 6 SEEOIVDFKEAFGLFDKGDGCTTVEELATVIRSLDQNTPEELQDMISEVDADGNGTIE 65

QY 123 LTAFLHNNGDHTVRLDRWEPELNEAIPNDERDPTTVAATLTKLL--TGELLTLASRQ 180
Db 66 FDEFLSLMAKKVKTODA--EELKEAFKVPDKQNGTIS--ASELRHVMINIGELTDEEVE 123

QY 181 QUIDMWEAD 189
Db 124 QMT--KEAD 130

RESULT 7

US-09-252-991A-29942
; Sequence 29942, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUCINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.116
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29942
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (116)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match 4.8%; Score 64.5; DB 4; Length 178;
Best Local Similarity 26.8%; Pred. No. 40;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

QY 110 ANLLTTITG---GPKELTAFL--LHNNGDHTVRLDRWEPELNEAIPNDERDPTTVAAMA 162
Db 25 ARVLLVELMOORRGEELAAHAGLAHHPGDHVAR-----AGEDLP---RDVQGLGAAA 74

QY 163 TTKRLTLTGELL-----TLASRQQLIDMMEADKYAGLRSALPAGFIADRS 211
Db 75 LRHLLYTAGALOGHEHREGIGDGTAAGEQAVVG--QDEYVAGPVV--GLQAGLFLAVQGD 130

QY 212 AGERGSRGIIAALGPDKR 229
Db 131 A-----LVGVGQAGK 141

QY 212 AGERGSRGIIAALGPDKR 229
Db 131 A-----LVGVGQAGK 141

RESULT 8
US-09-413-814-87
; Sequence 87, Application US/09413814
; Patent No. 625064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecher, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolysaccharide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 87
LENGTH: 198
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-87

Query Match 4.8% Score 64.5; DB 3; Length 198;

Best Local Similarity 22.4%; Pred. No. 48;
Matches 62; Conservative 30; Mismatches 78; Indels 107; Gaps 15;

QY 1 HPEVLVKKVDAEDQIARVGYIELDNLGKILSFSEFSEPPMSTF-----KVLICGAV 55
DB 7 MEADVAVDKTS--AAFG-----DAVHKVLEGGAGVETATYTFGEHFEPSVLCVAS 60
QY 56 LSRIDAGQEQOLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAITSMDNTANLLIT 115
DB 61 LVR-----MGVRIIARAA---TDRQAD--IIR 82
QY 116 TTGGKPELTAFLNMGDHTRLDWEPELNAIPNDEDTTPVAMATTLRKLTIGELLT 175
DB 83 AVG-----ATRVQLLETMEGRV---GADITMPLAQ-----DLID 114
QY 176 LASROOLIDMEADKVAGPLLRSLAPGWFIADK---SGAGER-----GSRGIIAIG 225
DB 115 LASHYVVPW---NAHGPLYGQTL-AGSKTRQIRIVLVGRPHTRNRPDKPRLEAFT 169
QY 226 PDGKPSRIVIVYTTGSGATM-----DERNPOIAEIG 256
DB 170 PD-----YVIRQDGTLLLVGDSDDVSHPVAEVG 197

RESULT 9

US-08-624-677A-2
Sequence 2, Application US/08624677A
Patent No. 6476192
GENERAL INFORMATION:
APPLICANT: Lally, Nicola C.
APPLICANT: Jenkins, Mark C.
APPLICANT: Dubey, Jitender P.
TITLE OF INVENTION: Antigens useful for the serodiagnosis of
TITLE OF INVENTION: Neosporosis
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Graeter, Janelle S.
STREET: Rm. 411, Bldg. 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624.677A
FILING DATE: 15-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0228.95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-504-6629
TELEFAX: 301-504-5060
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-624-677A-2

Query Match 4.8% Score 64; DB 4; Length 203;
Best Local Similarity 23.2%; Pred. No. 56;
Matches 38; Conservative 22; Mismatches 74; Indels 30; Gaps 6;

QY 54 AVLSRIDAGQEQOLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAA--ITMSDNTAAN 111
DB 33 AGVSNDGDDDAAGNPVD-----SDVTDAITDGEPRVYSGQPHHTQKSLIKK 83
QY 112 LLTTTGGKPELTAFLNMGDHTRLDWEPELNAIPNDEDTTPVAMATTLRKLTIGELLT 157
DB 84 LAVPVGA---LTSYL--VADRVLPDLTSAEGETESIPKRRKVTAGIALVAFAFA 138
QY 158 PVAMATTLRKLTIGELLTASROOLIDMEADKVAGPLLRSLAP 201
DB 139 GIGLAKTRHRYPKSKTYASEDSALGNSBEOYVEGTVNGSSDP 182

RESULT 10

US-08-225-480-4
Sequence 4, Application US/08225480
Patent No. 581793
GENERAL INFORMATION:
APPLICANT: LEVY, STUART B.
TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
TITLE OF INVENTION: ASSAYS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225.480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,085
FILING DATE: 28-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: T0359/7003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-225-480-4
Query Match 4.7% Score 63.5; DB 2; Length 144;
Best Local Similarity 29.5%; Pred. No. 38;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

QY 66 LGRRIHY--SQND--LVEY--SP--VTEKHLTDGMTVRELCSAITSMDNTANLLITIG 118
DB 14 LGRLIHYVQKQRRLINELYSPLDITAAQFVLSIR--CAACIT----- 56
QY 119 GKPELTAFLNMGDHTRLDWEPELNAIPNDEDTTPVAMATTLRK 167
DB 57 -PVELKRVSLVDGALTRMLDRLVCKGWERLPN--PDKRGVLYKLTGGNAICRCHQ 113
QY 168 LITGELLTASROOLIDMEADKVA--GPLLRSLAP 201

Db 114 LVGODL-----HQLTKRLTADEVAATLEYLLKKVLP 144

RESULT 11

US-09-118-445-4

Sequence 4, Application US/09118445

Patent No. 6391545

GENERAL INFORMATION:

APPLICANT: LEVY, STUART B.

TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON ASSAYS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/118,445

FILING DATE: 17-Jul-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/225,480

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: 70359/7003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-720-3500

TELEFAX: 617-720-2441

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 144 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-118-445-4

Query Match 4.7%; Score 63.5; DB 4; Length 144;

Best Local Similarity 29.5%; Pred. No. 38;

Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

QY 66 LGRIHY--SOND--LVEY--SP--VTEKHLTDMTVELCSAATMSDNFAANLLTTIG 118

Db 14 LGRILHMVQKKDRLNELYSPLDITAAQFKVLSIR--CAACT----- 56

QY 119 GPKELIAFLH-NMGHVTIRLDL-----WEPELNEAIPNDERD-----TTPPVAAATTLRK 167

Db 57 -PVELKVLVDGLALTRMLDLVCKGWVERLPN--PNDKRGVLVKLTGGAICEQCHQ 113

QY 168 LITGELLTLASROQLIDMWEADKVA--GPLLSALP 201

Db 114 LVGODL-----HQLTKRLTADEVAATLEYLLKKVLP 144

RESULT 12

US-09-252-991A-24737

Sequence 24737, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196,136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 24737

LENGTH: 160

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24737

Query Match 4.7%; Score 63.5; DB 4; Length 160;

Best Local Similarity 37.5%; Pred. No. 44;

Matches 18; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

QY 3 ETLVAVKDAEDQLGARVGIETDLSNGKITL-----ESFRSEERPPMS 45

Db 109 ELLEKVKSFLLKSGAIVLEVDENPNKAVSFYKSGFEFRERFVMS 156

RESULT 13

US-08-858-207A-519

Sequence 519, Application US/08858207A

Patent No. 6348328

GENERAL INFORMATION:

APPLICANT: Black, Michael

APPLICANT: Hodgson, John

APPLICANT: Knowles, David

APPLICANT: Nicholas, Richard

APPLICANT: Stodola, Robert

TITLE OF INVENTION: No. 6348328el Compounds

NUMBER OF SEQUENCES: 552

CORRESPONDENCE ADDRESS:

ADDRESSEE: SmitKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/017670

FILING DATE: 14-MAY-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimmil, Edward R.

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50475

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 519:

SEQUENCE CHARACTERISTICS:

LENGTH: 170 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6348328e

US-08-858-207A-519

Query Match 4.7%; Score 63.5; DB 4; Length 170;

Best Local Similarity 21.1%; Pred. No. 49;

Matches 45; Conservative 34; Mismatches 71; Indels 63; Gaps 8;

OY 15 LGARVGYELDLNSGILLESFRSEERFPMSTFKVLLCGAVLSRIDAG--EQLGRRHY 72
 DB 1 LRKNIGLVQDLISIRELSPISPTKKSVMRFK-----AAAFAVDASFIQELFORVD- 54
 OY 73 SONDVEYSPVTEKHILTDGTVRELCSAATMSDNTANLLTTIGGKELAFIHNMGD 132
 DB 55 -----SPVSEKSGSFSTGQROLAPARTVASQ-----PKLLI----- 86
 OY 133 HVRRLDRWEPELNEALPNDERDTTPVAMATTLRLKLLTGELLTLASROOLDMEADKVA 192
 DB 87 -----LDEATANIDSEESLV--QASLAKMRGR-TTIAIARLSTIQANCI- 131
 OY 193 GPLNSALPAGWFIADKSGAGERSGCIIALG 225
 DB 132 -----YLDKGRITESTHEELLALG 152

RESULT 14

US-09-252-991A-22681
 ; Sequence 22681, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 22681
 ; LENGTH: 132
 ; TYPE: PR
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-22681

Query Match 4.6%; Score 62; DB 4; Length 132;
 Best Local Similarity 29.8%; Pred. No. 46;
 Matches 28; Conservative 8; Mismatches 34; Indels 24; Gaps 4;

OY 144 LINEATPNDERDTTPVAMATTLRLKLLTGELLTLASROOLDMEADKVAGPLRSALP-- 201
 DB 34 LRQGIPT--HAGRPQATVLPMEPNCE-----RRVAFPMKSPPPPR 74
 OY 202 AGWFIADKSGAGERSGCIIALG--ALGPDGKPSR 232
 DB 75 APEYADRAAGPGPGCTTGCCAPRRAPAGVAPAR 108

RESULT 15

US-08-482-142-195
 ; Sequence 195, Application US/08482142
 ; Patent No. 5820862
 ; GENERAL INFORMATION:
 ; APPLICANT: Garman, Richard
 ; APPLICANT: Greenstein, Julia
 ; APPLICANT: Kuo, Mel-chang
 ; APPLICANT: Rogers, Bruce
 ; APPLICANT: Franzen, Henry
 ; APPLICANT: Chen, Xian
 ; APPLICANT: Evans, Sean
 ; APPLICANT: Shaked, Ze'ev
 ; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 ; NUMBER OF SEQUENCES: 207
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
 ; STREET: 610 LINCOLN STREET

CITY: WALTHAM
 STATE: MA
 COUNTRY: USA
 ZIP: 02154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/482,142
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/445,307
 FILING DATE: 07 June 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: CRAIG, ANNE I.
 REGISTRATION NUMBER: 32,976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 466-6000
 TELEFAX: (617) 466-6040
 INFORMATION FOR SEQ ID NO: 195:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 181 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-482-142-195

Query Match 4.6%; Score 62; DB 2; Length 181;
 Best Local Similarity 19.9%; Pred. No. 79;
 Matches 34; Conservative 25; Mismatches 66; Indels 46; Gaps 5;

OY 4 TLVYKDAEDQIGARYG--IELDLSGKILLESFRSEERFPMSTFKVLLCGAVLSRIDA 61
 DB 17 THAKIRDVATESAYLATRYNTSDLSQELVDEYIQHNGVQVESTYRY----- 65
 OY 62 GOELGRRIHYSQ-----NDLVEYSPVTEKHILTDGTVRELCSAATMSDNTANLL 114
 DB 66 AREOSCRNRNAOLLEAVEFNOMTKTKIEIKASIDGLEV----- 105
 OY 115 TTIGGKELAFIHNMGDHYTRLDRWEPELNEAIPND-----ERDTTPVAM 161
 DB 106 --IIGIKDDAFRHYDGRITIIQRDNGYQISGNAPAEIDLRQMRTVPIRM 154

Search completed: September 10, 2003, 12:31:04
 Job time: 11.8571 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:30:56 ; Search time 17.5714 Seconds

(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_37S_157T

Perfect score: 1345

Sequence: 1 HPETLVKVKADIEDQLGARVY.....TMDERNRQIAEIGASLIRKM 263

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters:

278642

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*

6: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubppaa/US08_PUB.pep:*

8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*

13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

17: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

18: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 457 | 34.0 | 94 | US-10-102-806-760 | Sequence 760, App |
| 2 | 75 | 5.6 | 202 | US-10-127-816-9 | Sequence 9, App1 |
| 3 | 75 | 5.6 | 202 | US-10-142-717-12 | Sequence 12, App1 |
| 4 | 70.5 | 5.2 | 193 | US-09-805-354-8 | Sequence 8, App1 |
| 5 | 70.5 | 5.2 | 193 | US-10-144-259-8 | Sequence 8, App1 |
| 6 | 70 | 5.2 | 206 | US-10-156-761-13867 | Sequence 13867, A |
| 7 | 69.5 | 5.2 | 184 | US-10-101-464A-765 | Sequence 765, App |
| 8 | 69 | 5.1 | 149 | US-10-156-761-8136 | Sequence 8136, App |
| 9 | 67.5 | 5.0 | 149 | US-10-233-926-4 | Sequence 4, App1 |
| 10 | 67.5 | 5.0 | 187 | US-09-948-018-8 | Sequence 8, App1 |
| 11 | 67 | 5.0 | 178 | US-09-738-626-5680 | Sequence 5680, App |
| 12 | 67 | 5.0 | 194 | US-10-156-761-8623 | Sequence 8623, App |
| 13 | 66 | 4.9 | 202 | US-10-127-816-11 | Sequence 11, App |
| 14 | 64.5 | 4.8 | 189 | US-09-815-242-5862 | Sequence 5862, App |
| 15 | 64.5 | 4.8 | 189 | US-09-815-242-12979 | Sequence 12979, A |

| | | | | | |
|----|------|-----|-----|---------------------|--------------------|
| 16 | 64.5 | 4.8 | 191 | US-10-156-761-8978 | Sequence 8978, App |
| 17 | 64 | 4.8 | 162 | US-09-738-626-4796 | Sequence 4796, App |
| 18 | 64 | 4.8 | 179 | US-09-764-868-757 | Sequence 757, App |
| 19 | 64 | 4.8 | 179 | US-10-106-698-4858 | Sequence 4858, App |
| 20 | 64 | 4.8 | 202 | US-10-106-698-4858 | Sequence 12, App1 |
| 21 | 63.5 | 4.7 | 88 | US-10-100-252-6 | Sequence 6, App1 |
| 22 | 63.5 | 4.7 | 144 | US-10-131-406-4 | Sequence 4, App1 |
| 23 | 63.5 | 4.7 | 202 | US-10-156-761-13460 | Sequence 13460, A |
| 24 | 63.5 | 4.7 | 206 | US-09-738-626-5425 | Sequence 5425, App |
| 25 | 62.5 | 4.6 | 160 | US-09-882-227-230 | Sequence 230, App |
| 26 | 62.5 | 4.6 | 190 | US-10-156-761-9507 | Sequence 9507, App |
| 27 | 62.5 | 4.6 | 191 | US-09-764-870-390 | Sequence 390, App |
| 28 | 62.5 | 4.6 | 191 | US-10-125-540-390 | Sequence 35777, A |
| 29 | 62 | 4.6 | 174 | US-09-864-761-35777 | Sequence 523, App |
| 30 | 62 | 4.6 | 189 | US-09-764-870-523 | Sequence 523, App |
| 31 | 62 | 4.6 | 189 | US-10-125-540-523 | Sequence 523, App |
| 32 | 62 | 4.6 | 191 | US-09-927-827-64 | Sequence 64, App1 |
| 33 | 61.5 | 4.6 | 142 | US-09-862-027-31 | Sequence 31, App1 |
| 34 | 61.5 | 4.6 | 196 | US-09-738-626-6624 | Sequence 6624, App |
| 35 | 61 | 4.5 | 185 | US-10-101-464A-616 | Sequence 616, App |
| 36 | 60.5 | 4.5 | 202 | US-10-189-346-16 | Sequence 16, App |
| 37 | 60 | 4.5 | 190 | US-09-738-626-4637 | Sequence 4637, App |
| 38 | 59.5 | 4.4 | 182 | US-10-156-761-12056 | Sequence 12056, A |
| 39 | 59.5 | 4.4 | 196 | US-10-219-220-265 | Sequence 265, App |
| 40 | 59.5 | 4.4 | 206 | US-09-741-669-461 | Sequence 461, App |
| 41 | 59.5 | 4.4 | 206 | US-09-815-242-10335 | Sequence 10335, A |
| 42 | 59 | 4.4 | 111 | US-09-867-550-1868 | Sequence 1868, App |
| 43 | 59 | 4.4 | 134 | US-09-768-235B-40 | Sequence 40, App1 |
| 44 | 59 | 4.4 | 180 | US-09-922-683-5 | Sequence 5, App1 |
| 45 | 59 | 4.4 | 181 | US-10-156-761-10043 | Sequence 10043, A |

ALIGNMENTS

RESULT 1

US-10-102-806-760

Sequence 760, Application US/10102806

Publication No. US20030054422A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103PICI

CURRENT APPLICATION NUMBER: US/10/102, 806

CURRENT FILING DATE: 2002-03-22

PRIOR APPLICATION NUMBER: 09/925, 298

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05881

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124, 270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 846

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 760

LENGTH: 94

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (80)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (91)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-102-806-760

Query Match

Best Local Similarity 34.0% Score 457; DB 15; Length 94;

Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

NDLVESPYVEKHITDGTAVELCSAATMSDNFANLITIGPKELTAPFHNMGDHY 134

NDLVESPYVEKHITDGTAVELCSAATMSDNFANLITIGPKELTAPFHNMGDHY 60

OY 135 TRLDREPELNEAIPNDERDTTPVAMATT 164
 DB 61 TRLDREPELNEAIPNDERDTTPVAMATT 90

RESULT 2

US-10-127-816-9
 ; Sequence 9, Application US/10127816
 ; Publication No. US20030104416A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sheppard, Paul O.
 ; APPLICANT: Fox, Brian A.
 ; APPLICANT: Klueber, Kevin M.
 ; APPLICANT: Taft, David W.
 ; APPLICANT: Kindsvogel, Wayne R.
 ; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
 ; FILE REFERENCE: 01-17
 ; CURRENT APPLICATION NUMBER: US/10/127,816
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/285,408
 ; PRIOR FILING DATE: 2001-04-20
 ; PRIOR APPLICATION NUMBER: US 60/286,482
 ; PRIOR FILING DATE: 2001-04-25
 ; PRIOR APPLICATION NUMBER: US 60/341,050
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 60/341,105
 ; PRIOR FILING DATE: 2001-10-22
 ; PRIOR APPLICATION NUMBER: US 09/895,834
 ; PRIOR FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/285,424
 ; NUMBER OF SEQ ID NOS: 59
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9
 ; LENGTH: 202
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-127-816-9

Query Match 5.6%; Score 75; DB 15; Length 202;
 Best Local Similarity 21.4%; Pred. No. 13;

Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEQLGRIRHYSQND--LVEYSPYTEKHL-----TDGMTVREL---- 97
 DB 17 LLLAAVLTRQADVPVPRATRLPVEAKDCHIQFSLSPRELQAFKAKDAIEKRLLEKDL 76
 OY 98 -CSA-----AITMS--DNTANLLTTIGCPRELTAFLHN 129
 DB 77 RCSSHLFPRAMD LKQLQVQERPKALQAEVALTLKVMENMTDSALATILGQPLHTLSHHS 136
 OY 130 MGDHYT-----RLDRPEPELNEAIPNDERDTTPVAMATTLLKLTGELLTL 176
 DB 137 QLQCTQQLQATAEPSPSRRLSRWLRLQLEA-QSKETPGCLEASVTSNLFRLTLRDLKCV 195
 OY 177 ASRQQLI 183
 DB 196 ANGDQCV 202

RESULT 3

US-10-142-717-12
 ; Sequence 12, Application US/10142717
 ; Publication No. US20030104579A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Immunex Corporation
 ; APPLICANT: Baum, Peter R.
 ; APPLICANT: Mosley, Bruce A.
 ; APPLICANT: Ketchum, Randall R.
 ; APPLICANT: Taylor, Scott L.
 ; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
 ; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717

; CURRENT FILING DATE: 2002-05-08

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 12

; LENGTH: 202

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-142-717-12

Query Match 5.6%; Score 75; DB 15; Length 202;
 Best Local Similarity 21.4%; Pred. No. 13;

Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEQLGRIRHYSQND--LVEYSPYTEKHL-----TDGMTVREL---- 97
 DB 17 LLLAAVLTRQADVPVPRATRLPVEAKDCHIQFSLSPRELQAFKAKDAIEKRLLEKDL 76
 OY 98 -CSA-----AITMS--DNTANLLTTIGCPRELTAFLHN 129
 DB 77 RCSSHLFPRAMD LKQLQVQERPKALQAEVALTLKVMENMTDSALATILGQPLHTLSHHS 136
 OY 130 MGDHYT-----RLDRPEPELNEAIPNDERDTTPVAMATTLLKLTGELLTL 176
 DB 137 QLQCTQQLQATAEPSPSRRLSRWLRLQLEA-QSKETPGCLEASVTSNLFRLTLRDLKCV 195
 OY 177 ASRQQLI 183
 DB 196 ANGDQCV 202

RESULT 4

US-09-805-354-8
 ; Sequence 8, Application US/09805354
 ; Publication No. US20030078375A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Atmadou, M. Amin
 ; APPLICANT: Li, Rui
 ; APPLICANT: Xiong, Jian-Ping
 ; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
 ; FILE REFERENCE: 00786-536001
 ; CURRENT APPLICATION NUMBER: US/09/805,354
 ; CURRENT FILING DATE: 2002-06-04
 ; PRIOR APPLICATION NUMBER: US 09/758,493
 ; PRIOR FILING DATE: 2001-01-11
 ; PRIOR APPLICATION NUMBER: US 60/221,950
 ; PRIOR FILING DATE: 2000-07-31
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 193
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-805-354-8

Query Match 5.2%; Score 70.5; DB 11; Length 193;
 Best Local Similarity 24.2%; Pred. No. 35;

Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;

OY 59 IDAGEQLGRIRHYSQNDLVEYSP--VTEKHLTDGMTVRELCSAAITMSDNTANLLTTT 116
 DB 36 IGPQQLQV-----VVQGEDVVEHFLNDVRSYKDVVEAASHIRQR----- 77
 OY 117 IGGPKELTAPLHNNMGHVTRLDRNE-----PLNEAIPNDERD 154
 DB 78 -GGTEFRTAF-----GIFARSEAFQKGRGKAKKVIIVTQESHSDSPLEEVYQOOSERD 132
 OY 155 TTPVAMA 162
 DB 133 NVTRYAVA 140

RESULT 5


```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8136
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-8136.

Query Match
Best Local Similarity 30.9%; Score 69; DB 15; Length 149;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;

QY 168 LTTGELLT-----ASROQLDMMDKVAAPLLRSALPAGWFIADKSG----- 211
DB 50 LLDIELTTIKRLIVASVDRAKEMGIDMWHED---PALSSHADGCRLEAEERRLREI 105
QY 212 AGERSRGITAAAGPDGKPSR 232
DB 106 AGLRDAALPSAEGPAEKPER 126

RESULT 9
US-10-233-926-4
; Sequence 4, Application US/10233926
; Publication No. US20030131382A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
; FILE REFERENCE: B81419 US NA
; CURRENT APPLICATION NUMBER: US/10/233,926
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: US/09/735,846
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
; US-10-233-926-4

Query Match
Best Local Similarity 21.4%; Score 67.5; DB 16; Length 149;
Matches 22; Conservative 23; Mismatches 33; Indels 25; Gaps 4;

QY 159 VAAATTLRKLL-----TCELLTASROQLDMMDKVAAPLLRSALPAGWFIADKSGA 212
DB 14 LSLSTSLPNTLAMDHNAEAAPQSSQEEEDMKKEAGDGDV-----EVAADRGGG 64
QY 213 GEGSRGITAAGPDGKPSRIV-----IYTGSAQATMDERNR 250
DB 65 GGAANGI-----PEGRRIRYVADGIYDLFHGHAKSLEDAKR 102

RESULT 10
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: Thell et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
```

```

; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-948-018-8

Query Match
Best Local Similarity 22.4%; Score 67.5; DB 10; Length 187;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

QY 112 LTTTGGPKETAFLNHGDHVRILDRWEPE-----LNEAIPNDEROTTPVAMAT- 163
DB 11 LLLMLAPPEAS-----QYCGRLPYWNPDKKCCSCLOREGP-----PPCRALETG 58
QY 164 -----TLRKLLTGEELLTLASR--QQLLDMMEA-----DKVAGP-----L 195
DB 59 DTWRKASLPLSLRSLSIASPLSLRLDELVELELIVLLDPEPPGCGMAHGTTRHIA 118
QY 196 LRSALPAGWFIADKSGAGERSGITAALGPDGKPSR--IYVITGSAQATMDERNROIA 253
DB 119 ARYGLPAW-----STFAVSLRPSRSLRALIEMVAREPSASLGQSTHIA 165
QY 254 ETG 256
DB 166 QLG 168

RESULT 11
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-5680

Query Match
Best Local Similarity 24.8%; Score 67; DB 10; Length 178;
Matches 25; Conservative 17; Mismatches 37; Indels 22; Gaps 4;

QY 171 GELLTASROQLDM-----MEADKVAAPLLRSALPAGWFIADKSGAGERSGITA 223
```

```
DB 61 GELFPAEORGEINFCAGTLEWSTPGVNDPL---TLPRHM-----RRNGRLVA 107
OY 224 ICPDGK--SRIVYITTSQATMDERNROIAETIGASLKH 262
DB 108 LDQCKRVARIGALNDAETHVLIERNKILEVTTLELAH 148

RESULT 12
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIRA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 194;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;

OY 202 AGFWIADKSGAGERSRG---TIALGPDG--KPSRIVITYTT 239
DB 75 SGFWYTPRSAGERTEGAEYRITTAIGLVAHEPVVAVVVT 118

RESULT 13
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11

Query Match
Best Local Similarity 4.9%; Score 66; DB 15; Length 202;
Matches 41; Conservative 23; Mismatches 65; Indels 62; Gaps 8;

OY 50 LLCGAVISRIDAGQDQLGRRIHYSND--LVESPYTERLU-----TDGMYREL----- 97
DB 17 LLAAVLTFTQADPPRATRLPEVAKDCHIAQPKSLPKELQAFKRAKGAIEKRLERDM 76
OY 98 -CSA-----ATMS--DNTAANLLTTTGGPKELTAFLN 129
DB 77 RCSSHLISRANLKOLOVERPKALOAEVALTKVMENINDSALTTLGQPLNLSHIHS 136
OY 130 MGDHYT-----RLDRWEELMEA---IPNDRDTTTPVAAATTLKLLTGE 172
DB 137 QLOTCTQLOATAPRPKPPSRRLSRWLRLQEAOSKETPGCLEDSVT-----SNLFOLLRD 191
OY 173 LLLTASROQLI 183
DB 192 LKCVASGDQCV 202

RESULT 14
US-09-815-242-5862
; Sequence 5862, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EXTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5862
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5862

Query Match
Best Local Similarity 4.8%; Score 64.5; DB 9; Length 169;
Matches 42; Conservative 32; Mismatches 84; Indels 47; Gaps 8;

OY 4 TLVKYDAEDQLCARGYIELDNSGKILSFSEERPPMSTFVLLGCA----- 54
DB 10 TLIKI--VOEDNQRTFKTELTKNIDQVVEWLNMOO-----IEKICLTGAGYIAENI 61
```

[illegible]

```
Search completed: September 10, 2003, 12:33:19
Job time : 17.5714 secs
```

```

US-09-815-242-12979
; Sequence 12979, Application US/09815242
; Patent No. US2002061569A1
;
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
;
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE OF INVENTION: Prokaryotes
;
; FILE REFERENCE: ELITRA.011A
;
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
;
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12979
;
; LENGTH: 189
;
; TYPE: PRF
;
; ORGANISM: Staphylococcus aureus
US-09-815-242-12979

```

| | | | | | | |
|----|--|------------------|--------------------|------------|-------------|--|
| | Query Match | 4.8% | Score 64.5; | DB 9; | Length 189; | |
| | Best Local Similarity | 20.5%; | Pred. No. 1.4e+02; | | | |
| | Matches 42; | Conservative 32; | Mismatches 84; | Indels 47; | Gaps 8 | |
| OY | 4 TLVKKVKAEDQGLGAVGYIELDNSGKILIESPSESRFPFMSFFKVLLCGA----- | 54 | | | | |
| Db | 10 TLIRI--VQEODNORTFKTELTKRIDVOYEMLNCOO-----TEKLCTGGNAGVIAENI | 61 | | | | |
| OY | 55 -----VLRIADAGDQLGRRIHYSONDVLEY-----SPVTEKHILTGDGMTVELSMAITM | 104 | | | | |
| Db | 62 NIPAOIFVEFDDASGSLGLEKEOGHDLDVYIPRANGTGTSLHTFDOSORRVGSGTGCG | 121 | | | | |
| OY | 105 SDNTAAANLTTTIGCPKETAEFLHNKGDVYTRIDREPELENEALPRNDERTTTPYAAMAT | 164 | | | | |
| Db | 122 GMIGLGAYLLSQSIDTYDKOLT-----DMAQNGDR-----NTIDLKVRHIYKDTPEPI---- | 167 | | | | |
| OY | 165 LRLKLTGELLTLASROQLIDMMHAD | 189 | | | | |
| Db | 168 -----PGD-LTPAANFGIVLHHNDAD | 186 | | | | |

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_37S_157T

Sequence: 1 HPETLVKVKDAEDQLGARVGT.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_76:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 39.0 | 105 | 2 | JC2566 |
| 2 | 77 | 5.7 | 191 | 2 | S67447 |
| 3 | 74.5 | 5.5 | 113 | 2 | T45195 |
| 4 | 74.5 | 5.5 | 172 | 2 | AD3606 |
| 5 | 72.5 | 5.4 | 184 | 2 | T21126 |
| 6 | 72.5 | 5.4 | 200 | 2 | G97064 |
| 7 | 71.5 | 5.3 | 152 | 2 | T36984 |
| 8 | 71.5 | 5.3 | 195 | 2 | T36975 |
| 9 | 71 | 5.3 | 153 | 2 | E81708 |
| 10 | 71 | 5.3 | 177 | 2 | A83753 |
| 11 | 70 | 5.3 | 192 | 2 | G83096 |
| 12 | 70 | 5.2 | 106 | 2 | B84217 |
| 13 | 70 | 5.2 | 170 | 2 | AB0192 |
| 14 | 69.5 | 5.2 | 167 | 2 | D67360 |
| 15 | 68.5 | 5.1 | 145 | 2 | F84251 |
| 16 | 68.5 | 5.1 | 190 | 2 | T09136 |
| 17 | 68.5 | 5.1 | 192 | 2 | A83587 |
| 18 | 68 | 5.1 | 149 | 2 | F71252 |
| 19 | 67.5 | 5.0 | 131 | 2 | AD2281 |
| 20 | 67 | 5.0 | 116 | 2 | G82906 |
| 21 | 66.5 | 4.9 | 177 | 2 | D90227 |
| 22 | 66.5 | 4.9 | 198 | 2 | D95285 |
| 23 | 66.5 | 4.9 | 198 | 2 | D95285 |
| 24 | 65.5 | 4.9 | 442 | 2 | F56978 |
| 25 | 65.5 | 4.9 | 181 | 2 | AB1902 |
| 26 | 65.5 | 4.9 | 195 | 2 | AE0623 |
| 27 | 65.5 | 4.9 | 197 | 2 | H90211 |
| 28 | 65 | 4.8 | 148 | 2 | E75283 |
| 29 | 65 | 4.8 | 150 | 2 | T08585 |

| | | | | | | |
|----|------|-----|-----|---|--------|--------------------|
| 30 | 65 | 4.8 | 160 | 1 | E69186 | conserved hypotnet |
| 31 | 65 | 4.8 | 168 | 2 | B75498 | conserved hypotnet |
| 32 | 65 | 4.8 | 177 | 1 | B43387 | nonstructural prot |
| 33 | 65 | 4.8 | 177 | 2 | J01931 | nonstructural prot |
| 34 | 65 | 4.8 | 180 | 2 | G70912 | hypothetical prote |
| 35 | 64.5 | 4.8 | 145 | 2 | H75262 | hypothetical prote |
| 36 | 64.5 | 4.8 | 151 | 2 | D81333 | probable protein-t |
| 37 | 64 | 4.8 | 113 | 2 | D70580 | hypothetical prote |
| 38 | 64 | 4.8 | 142 | 2 | A71000 | hypothetical prote |
| 39 | 64 | 4.8 | 161 | 2 | C71409 | probable RNA polym |
| 40 | 64 | 4.8 | 162 | 2 | AC0769 | probable acetyltra |
| 41 | 64 | 4.8 | 177 | 2 | T48420 | hypothetical prote |
| 42 | 64 | 4.8 | 178 | 2 | T40124 | outer surface prot |
| 43 | 64 | 4.8 | 191 | 2 | E95333 | hypothetical prote |
| 44 | 63.5 | 4.7 | 164 | 2 | H82336 | regulator of sigma |
| 45 | 63.5 | 4.7 | 166 | 2 | C90029 | hypothetical prote |

ALIGNMENTS

RESULT 1
JC2566
bla protein - Pseudomonas aeruginosa (fragment)
C:Species: Pseudomonas aeruginosa
C:Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996
C:Accession: JC2566
R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 148[128], 81-86, 1994
A>Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro
A:Reference number: JC2565
A>Note: due to a typographical error the volume number 148 appears as 128
A:Accession: JC2566
A:Molecule type: DNA
A:Residues: 1-105 <WES>
C:Genetics:
A:Gene: bla
C:Superfamily: beta-lactamase I

Query Match 39.0% Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 7.6e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLKRLTGLTSLASROQLIDMEADKAVGPIRLRGALPAGWPIADKSGSGERSRG 219
Db 2 AMATTLKRLTGLTSLASROQLIDMEADKAVGPIRLRGALPAGWPIADKSGSGERSRG 61
QY 220 IIAALGPDGKPSRIIVITTTGSOATMDERNRQIAEIGASLIKHW 263
Db 62 IIAALGPDGKPSRIIVITTTGSOATMDERNRQIAEIGASLIKHW 105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: T38062; S67447
R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996
A:Reference number: 221766
A:Accession: T38062
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <MC2>
A:Cross-references: EMBL:269944; NID:g1217974; PIDN:CAA93808.1; PID:g1217978; GSPDB:G
A:Experimental source: strain 972h-; cosmid c1f12
C:Genetics:
A:Gene: SPAC1F12.04c
A:Map position: 1
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c
Query Match 5.7%; Score 77; DB 2; Length 191;

[illegible]

```

RESULT 3
T45195
hypothetical protein u1756t [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-20000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: 216912
A:Accession: T45195
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

```

```

Query Match      5.5%; Score 74.5; DB 2; Length 113;
Best Local Similarity 28.6%; Pred. No. 30;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8

Oy 130 MGDDV--TRLDKWEPELN---EAPNDERDQ---TTPVAMATRLKLLTGELLTLASRQ 180
      ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1 MGGGICMERGRCRWGNTQCPLRVVGDESPITLDGRASPELIIT--NLSPTTMSPPPS 58

Oy 181 QLDQWMEA-KVAVAPLL-----RSALPGWFIADKSGAGERSGRIIALGPDQPSR 232
      : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 59 RDDDWEEPPALQGTAVFDATGDKRIIMPVGT-----GASTRGS-GIASLSPFPQAPR 112

```

```

RESULT 4
AD3606
molybdopterin biosynthesis mog protein [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AD3606
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Gottman, E.; Sekov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3606
A>Status: preliminary
A:Molecule type: DNA
A:Releases: 1-172 <RUR>
A:Cross-references: GB:AE008918; PIDN:PAL54015.1; PID:q17984966; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME110773
A:Map position: II
C:Superfamily: molybdenum cofactor biosynthesis protein B moab

Query Match      5.5%  Score 74.5;  DB 2;  Length 172;
Best Local Similarity 25.5%;  Pred. No. 53;
Matches 40;  Conservative 14;  Mismatches 42;  Indels 61;  Gaps 9;

```

[illegible]

RESULT 5
T21126
ADP-ribosylation factor homolog F19H8.3 [similarity] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 02-Feb-2001
C:Accession: T21126
R:Steward, C.
submitted to the EMBL Data Library, March 1997
Reference number: Z19379

A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Accession: F19H8.3
A:Molecule type: DNA
A:Residues: 1-184 <WILD>
A:Cross-references: EMBL:293378; PIDN:CAB07583.1; GSPDB:GN00020; CESP:F19H8.3
A:Experimental source: clone F19H8
C:Genetics:
A:Gene: CESP:F19H8.3
A:Map position: 2
A:Introns: 143/2
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)
F:90-95/Region: nucleotide-binding motif B
F:127-130/Region: GTP-binding NKXD motif
F:2/Modified site: myristylated amino end (gly) (in mature form) *status predicted

| Query Match | 5.43: | Score 72.5: | DB 2: | Length 184: |
|-----------------------|---|------------------|----------------|---------------------|
| Best Local Similarity | 21.6%: | Pred. No. 86: | | |
| Matches | 46: | Conservative 38: | Mismatches 80: | Indels 49: Gaps 11: |
| QY | 11ESFSSEERFPMMSTFKVLLGAVLSRIDAGEQGLRRIRHYSQNDLVEXSPVTEKHLTD | 90 | | |
| DB | 6 VLKSPFS-----PSGREIRILLGL-----DNAGKTTILKGL--SSEYQVHTVP-----TK | 49 | | |
| QY | GMVYRELCSAAITWSDNTAAAILLTTIGCPKELATLHNKGHW-----TRLD | 138 | | |
| DB | 50 GFNVKTVAA---MGD---IRLNVMDIGGORSIRPWSNSYNEINDLLIFVIDSDNDKRFED | 102 | | |
| QY | 139 RWEPELNEAIPNDERDFTTPVAMATTIRKLITGELLITLASROOLDIMWEADKVAGPLLRS | 198 | | |
| DB | 103 ENMIEICELL-DEKRLKRVPLVILFNKQ-----DLVTAASSEETRLIND-----LLRD | 151 | | |
| QY | 199 ALPAGWFIADKSGAGERGSGITIAALGPDGKRS | 231 | | |
| DB | 152 RT---WHIOACSAALNKEGINDGITYWASNIKPA | 181 | | |

RESULT 6
G97064
spore coat protein CORUC [Imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R: Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97064

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <RUR>
A:Cross-references: GB:AE00437; PIDN:AAK79306.1; PID:g15024270; GSPDB:GMO0168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1338

| | | | | | | | |
|-----------------|-------|--------------|------|------------|----|--------|-----|
| Query Match | 5.4% | Score | 72.5 | DB | 2 | Length | 200 |
| Best Similarity | 25.0% | Pred. NO. | 96 | | | | |
| Matches | 35 | Conservative | 27 | Mismatches | 55 | Indels | 23 |
| | | | | | | Gaps | 7 |

```

0y 1 HPE TLVKVKAEDQLGARVGYIEDLNSGRILESF-SEERFPMMSTRKVLCCGLAVLSRI 59
Db 10 HP--VKIKKPNMAQLARVYITQYGGPDELAAISIRLSRFSWNP-----QATATLN 59
0y 60 DAGQEQLGRRIHYSQNDLVEYSPTVEKHLTDGATVRLCSAALT--MADNTANILITTT 117
Db 60 DIGTEELAH-----LETVGSIVRQLSRGISVELEKRSGLDAYFADHDSAIYPAASAA 110
0y 118 GGPKELTAFLLHNMGDHWTRL 137
Db 111 GNP-FTAAITQSKGDPITDL 129

```

RESULT 7
T36984 /
hypothetical protein SCU1.13 - Streptomyces coelicolor

C: Update: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
C: Accession: T36984
R: Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream,
submitted to the EMBL Data Library, August 1999
A: Reference number: 221618
A: Accession: T36984
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-112 <OLI>
A: Cross-references: EMBL:AL109949; PIDN:CAB52898.1; GSPDB:GN00070; SCOEDB:SCJ1.13
A: Experimental source: strain A3(2)
A: Genetics:
A: Gene: SCOEDB:SCJ1.13
C: Superfamily: Streptomyces coelicolor hypothetical protein SCJ9A.17

Query Match 5.38; Score 71.5; DB 2; length 152;
Best Local Similarity 23.48; Pred. NO. 80;
Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;

RESULT 8
T36975
hypothetical protein SCJ11.04 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
R:Accession: T36975
C:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T36975
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA

A:Residues: 1-195 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CNB52889.1; GSPDB:GN00070; SCOEDB:SCJ11.04
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.04

| | | | | |
|-----------------------|-----------------|-------------------|-----------|------------|
| Query Match | 5.3% | Score 71.5 | DB 2 | Length 195 |
| Best Local Similarity | 17.6% | Pred. No. 1.1e+02 | | |
| Matches 34 | Conservative 35 | Mismatches 63 | Indels 61 | Gaps 7 |

```
OY 56 LSRIDAGQEOGLRRIHYSQNLDVEISPVTEKHILTOGTVRELCSAITMSDNTAANLLT 115
      : : : : : | : : : : : | : : : : : | : : : : : | :
Db 46 VAALGEIEIDRRV-----IARELVEALVAASAEITGVTAEBEGETALVP 93
OY 116 TIGGPPELTAFLNMGDHVTRLDRMEPELNEAI--PNDERDPTTPVAMATTLRLKITTEL 173
      | : : : : : | : : : : : | : : : : : | : : : : : | :
Db 94 APASAEPGA-----VFHHQEGLSVSIVLSPNNOR-----I 124
OY 174 LFLASRQQLIDMWEADKYAGPLLRSLPA-----GMFTRDKSGAGERGSR 218
      | : : : : : | : : : : : | : : : : : | : : : : : | :
Db 125 LNVLDGRPGLEPRVAKDIATAALGITAAAKAVEGPRPKRLAERGWLIOEASCAFSAGR 184
OY 219 GIITALGPDGKPS 231
      : : : : : | : : : : : | : : : : : | : : : : : | :
Db 185 -LVAS--PGGDPS 194
```

RESULT 9
E81708

conserved hypothetical protein TC0378 (imported) - Chlamydia muridarum (strain Nigg)
C/Species: Chlamydia muridarum, Chlamydia trachomatis MOpn
C/Date: 31-Mar-2000 #sequence_revistion 31-Mar-2000 #text_change 11-May-2000
C/Accession: E81708
R/Read: T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gyllm, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39
A/Reference number: AB1500, M01D:20150255, PMID:10664935
A/Accession: E81708
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-153 <ET>
A/Cross-references: GB:AE002305; GB:AE002160; NID:g7190418; PIDN:AAF39236.1; PID:g71919
A/Experimental source: strain Nigg (MOpn)
C/Genetics:
/Gene: TC0378

| | | | | |
|-----------------------|-------|---------------|-------|----------------|
| Query Match | 5.3% | Score 71; | DB 2; | Length 153; |
| Best Local Similarity | 34.6% | Pred. No. 89; | | |
| Matches | 18; | Conservative | 10; | Mismatches 16; |
| | | | | Indels 8; |
| | | | | Gaps 22; |

```

QY      124 TAFLNMGDHYRKLDRWEP--ELNEAIPADERDITYTPVAMATTLKLLTGEI 173
        | : ||| | : : ||| : :: || | : ||| |
Db      101 TIVTAERGEHYTLLEFPPTNTDLNELMGDQKTPP-----RAEMLSGFL 146

```

```

RESULT 10
A:83753
hypothetical protein BH0825 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: AB3753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: AB3650; MIMD:20512582; PMID:11058132
A:Accession: AB3753
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-177 <SNO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:910173176; PIDN:BAE04544.1; GSPDB:G
A:Experimental source: strain C-125
A:Genetics:

```

A:Gene: BH0825
C:Superfamily: Bacillus subtilis hypothetical protein ykka

Query Match 5.3%; Score 71; DB 2; Length 177;
Best Local Similarity 20.3%; Pred. No. 1.1e+02;

Matches 38; Conservative 28; Mismatches 57; Indels 64; Gaps 8;

QY 42 PMWSTKYLGVLSRIDAGQDGLGRIRHYSNDLVEVSPYTEKHLTDGMTVRELCSAA 101
DB 11 PNDTSTVGLFYAMVEEND-----RLH-----HLIDVTEELLYKKG 47
QY 102 ITWSDNTAANLLTTIGPKELTAFLHMGDHYRLD-RWEPEL-NEAIPND----- 151
DB 48 SSGDENSEMOLL-----NHLTYVDVRYRWYRIRIGELPDSLEAEHGP 89
QY 152 -ERDTPPVAMATTLRLKLLTGELLTLASRQ-----LIDMEADRYVAGPLLSALP 201
DB 90 VDDKGRIPVYTSVQGLIERQRYVALIKETCALHDDILARWIPYEEERQATIRWGL- 148
QY 202 AGWFIAD 208
DB 149 --WMMAD 153

RESULT 11

GB3096 conserved hypothetical protein PA4399 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001

C:Accession: GB3096

R:Stover, C.K.; Pham, X.Q.; Erylan, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.

Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: GB3096

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-192 <STO>

A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:ANG07787.1; GSPDB:GN001

C:Genetics:

A:Gene: PA4399

C:Superfamily: conserved hypothetical protein DRI638

Query Match 5.3%; Score 71; DB 2; Length 192;
Best Local Similarity 26.0%; Pred. No. 1.2e+02;

Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

QY 53 GAVLSRI-----DAGQEL--GRIRHYSNDLVEVSPYTEKHLTDGMTVRELCSA----- 100
DB 2 GNLRSKIYTRTGDBGEGCLAGGRYPKSHPRIEAIGAVELNSQLLAELEAARGAHP 61
QY 101 -----ATWSDNTAANLLTTIGPKELTAFLHMGDHYRLD-----DWEPEL-----N 145
DB 62 GLEIYQALAPVQHR-----LFDIGELAMPYRALDETVARLESCIDIMWDELGLKN 116
QY 146 EAIPNDRDTPPVAMATTLRLKLLTGELLTLASRQQLIDMEADRYVAGPLLSALPAGWP 205
DB 117 FILPGSR-----PVAQAHVCRSLAR-----SARRCOALDQEEITLEGVGLYRLNLSLLF 168
QY 206 IADKSGAGRGSRGII--AALGPD 227
DB 169 VAAARAIAROGVAEILWEAARPD 192

RESULT 12

EB4217 conserved hypothetical protein Vng0594h [Imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: EB4217

R:Ng, M.V.; Kennedy, S.P.; Manalras, G.G.; Bergulst, B.; Pan, M.; Shukla, H.D.; Laeky
: Jellhauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja

Jung, K.H.; Alam, M.; Freltas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: AB4160; MUID:20504483; PMID:11016950

A:Accession: EB4217

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-108 <STO>

A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA619105.1; GSPDB:GN00138

C:Genetics:

A:Gene: VNG0594H

Query Match 5.2%; Score 70; DB 2; Length 108;
Best Local Similarity 27.5%; Pred. No. 67;

Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

QY 44 MSTFKYLCGAVLSRIDAGQDGLGRIRHYSNDLVEVSPYTEKHLTDGMTVRELCSAAT 103
DB 1 MDTVTYELDGLALRLAETD-----RVEFVRFDALEVDVTLRFRRHDXDRV-----GSIV 50
QY 104 MSDNTAANLLTTIGPKEL-----TAFLHMGDHYRLDR 139
DB 51 NDDGTDRMARTLVPGDSDFAVEPTSEVAIVDAATRTDR 92

RESULT 13

AB0192 conserved hypothetical protein YP01575 [Imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002

C:Accession: AB0192

R:Parhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M

deno-Farrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G

ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AB0192

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-170 <EUR>

A:Cross-references: GB:AL590842; PIDN:CAC90397.1; PID:g15979615; GSPDB:GN00175

C:Genetics:

A:Gene: YP01575

C:Superfamily: Escherichia coli hypothetical protein b0354

Query Match 5.2%; Score 70; DB 2; Length 170;
Best Local Similarity 21.0%; Pred. No. 1.2e+02;

Matches 22; Conservative 25; Mismatches 44; Indels 14; Gaps 3;

QY 56 LSRIDAGQELGRIRHYSNDLVEVSPYTEKHLTDGMTVRELCSAATNSDNTAANLLLT 115
DB 70 MNKIDISKGNIG--FNFTDNMLI-----KRIYVDKILNOAQLISGLAIA-----RLVV 115
QY 116 TIGPKELTAFLHMGDHYRLDRWEPELNEAIPNDESDTTPVA 160
DB 116 DNGSESYAIIIPASVADKIRDAASSIVLSNSLQSEEDGEDPYA 160

RESULT 14

DB7360 conserved hypothetical protein CC0895 [Imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: DB7360

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Elsen, J.; Heidelberg,

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Nariakleeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SP02_37S_157T

Perfect score: 1345

Sequence: 1 HPETLVKVAEDDGLGARVG.....TMDERNQIAETGSLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 77 | 5.7 | 191 | YDA4_SCHPO | Q10346 schizosacch |
| 2 | 75 | 5.6 | 201 | RACG_DICDI | Q9960 dictyostell |
| 3 | 68.5 | 5.1 | 196 | RAC2_LOTJA | Q40220 lotus japon |
| 4 | 68 | 5.1 | 122 | HCL3_THIFE | P80509 thlobacilli |
| 5 | 68 | 5.1 | 149 | NDK_TREPA | Q83974 treponema p |
| 6 | 66.5 | 4.9 | 116 | KBYA_GREPA | Q9PQH0 ureaplasma |
| 7 | 65.5 | 4.9 | 206 | KTHY_METAC | Q8H59 methanosarc |
| 8 | 65 | 4.8 | 177 | VNSC_RINDK | P35948 rinderpest |
| 9 | 65 | 4.8 | 177 | VNSC_RINDK | Q03339 rinderpest |
| 10 | 63.5 | 4.7 | 144 | MARR_ECOLI | P27243 escherichia |
| 11 | 63.5 | 4.7 | 185 | YCCG_ECOLI | P38522 escherichia |
| 12 | 63.5 | 4.7 | 196 | ALKH_BACSU | P50846 b khyg/kbpy |
| 13 | 63 | 4.7 | 176 | HSLV_THEMA | Q9WY21 thermotoga |
| 14 | 62.5 | 4.6 | 146 | HGB_RABIT | P02039 oryctolagus |
| 15 | 62.5 | 4.6 | 150 | PDUV_SALT | Q9XADM salmonella |
| 16 | 62.5 | 4.6 | 160 | TATB_HELPY | Q25700 helicobacte |
| 17 | 62.5 | 4.6 | 172 | YDEJ_ECOLI | P31131 escherichia |
| 18 | 62.5 | 4.6 | 182 | PYRE_STRCO | Q9X8T7 streptococ |
| 19 | 62.5 | 4.6 | 184 | HRPL_PSESY | P37929 pseudomonas |
| 20 | 62.5 | 4.6 | 195 | TRPF_THEVO | Q979V6 thermoplas |
| 21 | 62.5 | 4.6 | 200 | KRA_PELNE | Q9FSA9 pellia nees |
| 22 | 62 | 4.6 | 173 | Y265_BORBU | O51280 borrelia bu |
| 23 | 62 | 4.6 | 174 | IHBH_RAT | P17491 rattus norv |
| 24 | 62 | 4.6 | 182 | RRA_BELCH | O19990 belamanda |
| 25 | 62 | 4.6 | 182 | Y661_BIFLO | O89550 bifidobacte |
| 26 | 62 | 4.6 | 186 | RRR_RICPR | Q9Z6B rickettsia |
| 27 | 62 | 4.6 | 186 | YCEB_SALT | P40832 salmonella |
| 28 | 61.5 | 4.6 | 178 | HSLV_RALSO | O8Y347 raietonia s |
| 29 | 61.5 | 4.6 | 193 | SAR2_LYCES | P52884 lycopersico |
| 30 | 61 | 4.5 | 121 | SECR_HUMAN | P09683 homo sapien |
| 31 | 61 | 4.5 | 184 | ARL2_DROME | O06849 drosophila |
| 32 | 61 | 4.5 | 194 | RRA2_IITEN | O02824 iris ensata |
| 33 | 61 | 4.5 | 200 | TATB_CAUCR | Q9a6t1 caulobacter |

| | | | | | |
|----|------|-----|-----|------------|--------------------|
| 34 | 60.5 | 4.5 | 103 | RS10_NEIGO | P48851 neisseria g |
| 35 | 60.5 | 4.5 | 152 | YU33_YERPE | Q8ZCF8 yersinia pe |
| 36 | 60.5 | 4.5 | 156 | BFR_AZOVI | P22759 azotobacter |
| 37 | 60.5 | 4.5 | 160 | TATB_HELPY | Q9ZM58 helicobacte |
| 38 | 60.5 | 4.5 | 184 | MLR1_SCHPO | Q9UUG5 schizosacch |
| 39 | 60 | 4.5 | 148 | CALM_BLAEM | Q9HFY6 blastoclad |
| 40 | 60 | 4.5 | 161 | PIN_BPT4 | P07068 bacterioph |
| 41 | 60 | 4.5 | 178 | PYRE_ARCFU | Q28533 archaeoglob |
| 42 | 60 | 4.5 | 178 | UCRI_ANASP | P70738 anabaena sp |
| 43 | 59.5 | 4.4 | 103 | RS10_NEIMA | Q9J121 neisseria m |
| 44 | 59.5 | 4.4 | 135 | CCRN_PAROL | O57312 paracitichy |
| 45 | 59.5 | 4.4 | 196 | I196_ASFB7 | P27943 african swi |

ALIGNMENTS

RESULT 1
YDA4_SCHPO STANDARD: PRT: 191 AA.
AC Q10346;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1P12.04c in chromosome I.
GN SPAC1P12.04c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream K.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Haidago J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens S., Warren T., Whitehead S.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wandut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Jelaire V., Mortier S.,
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerretti L., Lowe T., McComble W.R., Paulsen I., Potashkin J.,
RA Sipkavski G.V., Ussery D., Barrell B.G., Nurse P.;
RT The genome sequence of Schizosaccharomyces pombe.;
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z69944; CAA93808.1; -;
DR PIR: T38062; S67447;
DR GeneDB_Spombe; SPAC1P12.04c; -;

KW Hypothetical protein.
 SO SEQUENCE 191 AA; 21549 MW; 65555347F0EBED16 CRC64;
 Query Match 5.7%; Score 77; DB 1; Length 191;
 Best Local Similarity 22.6%; Pred. No. 26;
 Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;
 Y 104 MSDNTAANLLTTIGPKELTAFLHMGDVTFLDRKEPELEAIPNDEROTTTVAAMAT 163
 B 1 MSYHSLMLQNPSSGIDKIAALVN-----VARLD-----PASSKSTAOQLVSMLN 46
 Y 164 TLRLTGLLTLASRQOLIDHMEADKAVAGPLRLSALPAGWFIADKSGAGERSGRTIAA 223
 B 47 EPRC-----ILRLPLGLKLVNFRKDSPEPTYSNAINIGYVTE--GLAFLGCKQIIISI 99
 Y 224 LGPDKP-----SRIVYITGSGATDNERNRQI----- 252
 B 100 ----SKPLEDKMLMSRFLMDTLTLTYQLLREKTEDEKRNQDLASNLASLPICLHKS 155
 Y 253 AEGASLIRK 262
 B 156 VENGAGLHKH 165
 RESULT 2
 RAC2_DICD1 STANDARD; PRT; 201 AA.
 AC Q9GSP0:
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RAS-related protein racg.
 GN RACG.
 OS Dictyostelium discoideum (Slime mold).
 CC Eukaryota; Eukaryota; Dictyostelidia; Dictyostelium.
 OX NCBI_Taxid=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RX MEDLINE-21127961; PubMed-11222756;
 RA Rivero F., Dalslich H., Glocker G., Noegel A.A.;
 RL "The Dictyostelium discoideum family of Rho-related proteins."; Nucleic Acids Res. 29:1068-1079(2001).
 CC -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF310893; AAC45130.1; -
 DR HSSP: P21181; IAM4.
 DR DictyDB: DD77777; racg.
 DR InterPro: IPR003578; GTPase_Rho.
 DR InterPro: IPR001806; Ras_transfmg.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam: PF00071; ras; 1.
 DR PRINTS: PRO0449; RASTRNSFRMG.
 DR SMART: SM00174; RHO; 1.
 DR TIGRfam: TIGR00231; small_gtp; 1.
 KW GTP-binding; Prenylation; Lipoprotein.
 FT NP_BIND 10 17 GTP (BY SIMILARITY).
 FT NP_BIND 57 61 GTP (BY SIMILARITY).
 FT NP_BIND 115 118 GTP (BY SIMILARITY).
 FT DOMAIN 32 40 EFFECTOR REGION (POTENTIAL).
 FT LIPID 158 198 GERANYL-GERANYL (BY SIMILARITY).
 SO SEQUENCE 201 AA; 22813 MW; F6CE586497CC169 CRC64;
 Query Match 5.6%; Score 75; DB 1; Length 201;
 Best Local Similarity 22.6%; Pred. No. 41;

Matches 36; Conservative 20; Mismatches 59; Indels 44; Gaps 7;
 Y 44 MSTFVLLCGAVLSRIDAGQEQLRRIHNSQNDLV-EYSP-VYERKHLTDGTFVELCSNA 101
 B 1 MRSIRVCVVG-----EGDIGKTSMLSTYSNSISNEYOPTVDNYSF----- 42
 Y 102 ITMSDNFAANLLTTIGPKELTAFLHMGDVTFLDRKEPELEAIPNDEROTTTVAAMAT 145
 B 43 LHMNRKPRYNLSMTDAGQEQFSKRLRLSTPQDVFLLCSLNPSSFSNLDSSVQELN 102
 Y 146 EAIPNDERDTTPVAMATTLRLKLTGELL--TLASRQOL 182
 B 103 ENCPN-----TPIVLVGTQMDLKSNSVILDRICEKQQL 135
 RESULT 3
 RAC2_LOTJA STANDARD; PRT; 196 AA.
 AC Q40220:
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE RAC-1like GTP binding protein RAC2.
 GN RAC2.
 OS Lotus japonicus.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae; Lotus.
 OX NCBI_Taxid=34305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-cv, Glu / B-129; TISSUE-Root nodules;
 RX MEDLINE-97231679; PubMed-9076991;
 RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;
 RL "Identification of new protein species among 33 different small GTP-binding proteins encoded by cDNAs from Lotus japonicus, and expression of corresponding mRNAs in developing root nodules."; Plant J. 11:237-250(1997).
 CC -1 SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Z73962; CAA98190.1; -
 DR HSSP: P21181; IAM4.
 DR InterPro: IPR003578; GTPase_Rho.
 DR InterPro: IPR001806; Ras_transfmg.
 DR InterPro: IPR005225; Small_GTP.
 DR Pfam: PF00071; ras; 1.
 DR PRINTS: PRO0449; RASTRNSFRMG.
 DR SMART: SM00174; RHO; 1.
 DR TIGRfam: TIGR00231; small_gtp; 1.
 KW GTP-binding; Prenylation; Lipoprotein.
 FT NP_BIND 13 20 GTP (BY SIMILARITY).
 FT NP_BIND 60 64 GTP (BY SIMILARITY).
 FT NP_BIND 118 121 GTP (BY SIMILARITY).
 FT DOMAIN 35 43 EFFECTOR REGION (POTENTIAL).
 FT LIPID 193 193 GERANYL-GERANYL (POTENTIAL).
 SO SEQUENCE 196 AA; 21704 MW; 8BB0A70C4828CDS CRC64;
 Query Match 5.1%; Score 68.5; DB 1; Length 196;
 Best Local Similarity 24.1%; Pred. No. 1.3e+02;
 Matches 45; Conservative 23; Mismatches 64; Indels 55; Gaps 11;
 Y 44 MSTFVLLCGAVLSRIDAGQEQLRRIHNSQNDLV-EYSP-VYERKHLTDGTFVELCSNA 99
 B 1 MSTAFIRK-----VTVGGAAGVAGKTCMLISTYNTPTPTDVP-----TYPDNS 44

QY 100 AATMSDNTANLLTTIGPKK-----LTAF-LHMGDHVTRIDNEPE 143
 DB 45 ANVVV-DGTVNIGLWDTAGQEDYNRLPLSYRGADVFLARSLSRASYENISKWIPE 103
 QY 144 LINEAIPNDERDTPVAMATLRLKLTGELLTL-ASRQOLIMMEADKVA---GPLLRS 199
 DB 104 LNHVAV-----IVPI-----VLVGTKLIDREDRQIPLDHPGATPTTAAAGEELKKA 149
 QY 200 LPAQWFI 206
 DB 150 IGAAYVL 156

RESULT 4
 HC3L_THIFE
 ID HC3L_THIFE STANDARD; PRT; 122 AA.
 AC P80509;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cytochrome-C3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
 DE (Fragments).
 GN HOXC.
 OS Thioacillus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 OX NCBI_Taxid=920;
 RN [1]
 RP SEQUENCE.
 RC STRAIN-ATCC 19859;
 RA MEDLINE-96641862; PubMed-8661919;
 RA Fischer J., Quentmeyer A., Kostka S., Kraft R., Friedrich C.G.;
 RT "Purification and characterization of the hydrogenase from
 Thioacillus ferrooxidans."
 RL Arch. Microbiol. 165:289-296(1996).
 CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 - 2 H(+) + 2
 CC ferrocyclochrome c3.
 CC -1- COFACTOR: IRON.
 CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
 CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
 CC KW Oxidoreductase; Iron.
 FT NON_CONS 20 21
 FT NON_CONS 29 30
 FT NON_CONS 35 36
 FT NON_CONS 42 43
 FT NON_CONS 59 60
 FT NON_CONS 72 73
 FT NON_CONS 78 79
 FT NON_CONS 87 88
 FT NON_CONS 98 99
 FT NON_CONS 107 108
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 13146 MW; EB4FAF365348DAOB CRC64;
 Query Match 5.1%; Score 68; DB 1; Length 122;
 Best Local Similarity 26.4%; Pred. No. 79;
 Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;
 QY 19 VGYIELDINSGILFSEFSEPPMSTFKVLLCGAVLSRIDAGQEQDGRRIHYSQNDLV 78
 DB 12 VGRVGGDDXSILEYFRN-----AILARFGG-----LG 41
 QY 79 EYSPVTEKHLTDGMVRELCSAATMSDNTANLLTTIGPKKELT 124
 DB 42 KTAAPTGTNVEIGV-----ISGDKDPQAGLVVT-----PREST 75

RESULT 5
 NDK_TREPA
 ID NDK_TREPA STANDARD; PRT; 149 AA.
 AC 083974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
 DE (Nucleoside-2-P kinase).
 GN NDK OR TP1010.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_Taxid=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RA MEDLINE-98332770; PubMed-9665876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton C.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardam J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterlinden T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 spirochete."
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
 CC OTHER THAN ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +
 CC nucleoside triphosphate.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the NDK family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AE001269; AAC65961.1; -.
 DR PIR: F71252; F71252.
 DR HSSP: P22887; INPK.
 DR TIGR: TP1010; -.
 DR HAMAP: MF_00451; -; 1.
 DR InterPro: IPR001564; NDK.
 DR Pfam: PF00334; NDK, 1.
 DR PRINTS: PRO1243; NUCDPKINASE.
 DR PRODOM: PD001018; NDK; 1.
 DR SMART: SM00562; NDK; 1.
 DR PROSITE: PS00469; NDP_KINASES; 1.
 DR KATransferase; Kinase; ATP-binding; Complete proteome.
 KW ACT_SITE 117 117
 KW TRANSFERASE; Kinase; ATP-binding; Complete proteome.
 SQ SEQUENCE 149 AA; 16657 MW; BB329539671E80BE CRC64;

Query Match 5.1%; Score 68; DB 1; Length 149;
 Best Local Similarity 21.6%; Pred. No. 1e+02;
 Matches 35; Conservative 25; Mismatches 64; Indels 38; Gaps 7;

QY 29 GKLESPFSEPPMSTFKVLLCGAVLSRIDAGQEQDGRRIHYSQNDLVYSVTEKHL 88
 DB 21 GEVLSRF---ERKGLVLTALRLC-----VDPAVEL---HYAEERKPPFSLIAYI 67
 QY 89 TDGMTVELCSAATMSDNTANLLTTIGPKKELTAFLEHMGDHTVRLDRMEPELNAI 148
 DB 68 TSAVV-----ALAFKGEANISIVRTLCG-----TRVEAOP---GTI 103
 QY 149 PND-ERDTTPVAMATLRLKLTGELLTLASRQOLIDMMEAD 189
 DB 104 RGFALRTTIVIVASDPSANRELALYFSAQDFVEMRDGN 145

RESULT 6
 RBEA_UREPA
 ID RBEA_UREPA STANDARD; PRT; 116 AA.
 AC 09P0H0;

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR U0321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Letkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Caswell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RT Urealyticum.";
RL Mature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE002129; AAF30730.1; -
DR HAMAP: MF_00003; -; 1.
DR InterPro: IPR000238; Rib_bind_facta.
DR Pfam: PF02033; RBFA; 1.
DR ProDom: PD007327; Rib_bind_facta; 1.
DR TIGRFAMs: TIGR00082; rbfA; 1.
DR PROSITE: PS01319; RBFA; FALSE_NEG.
KW rRNA processing; Complete proteome.
SQ SEQUENCE 116 AA; 13347 MW; A79DC71F0547514 CRC64;

Query Match 4.9%; Score 66.5; DB 1; Length 116;
Best Local Similarity 28.9%; Pred. No. 98;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

QY 144 LNAIAPDERDRTTPVAMATTLRLKLTGEL-----LTLASROOLDWME-ADKVAAPLL 196
DB 18 INNALAEINDKIARLAVTAVR--LSNDLSVAKITFLDAHKRESMLKVLNVKVSQ-L 74
QY 197 RSALPAGW-----FIADKS 210
DB 75 RSKLAEMTSTYKVPFLRFVIDET 97

RESULT 7
KTHY_METAC
ID KTHY_METAC STANDARD: PRT; 206 AA.
AC 08THS9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
GN TMK OR M4433.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., MacDonald P.,

RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McEwen R., Talmas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate -> ADP + thymidine
CC 5'-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE011164; AAM07774.1; -
DR HAMAP: MF_00165; -; 1.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin; 1.
DR TIGRFAMs: TIGR00041; dTMP_kinase; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
KW transferase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 23303 MW; A19C259C85423B3 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 206;
Best Local Similarity 18.7%; Pred. No. 2,4e+02;
Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;

QY 56 LSRID-AGORQLGRIRIYSONDLVEYSPTEKHILDGKMTVRELCSAATMSDNTANLL 114
DB 8 LEGIDSSGKSTVAKKD-ENSELVEFEVFTREPFG-----TLTGNVAVNAIO 55
QY 115 TTIGPKRELTAFLHNMDDHYTRLDWEPELNEAIPNDERDRTTPVAMATTLRLKLTGELL 174
DB 56 SDTDLAELEFLPADHAEHLAKLVPALLEDKTVSDRYSDRYAVOGTTLKRLDNP 114
QY 175 TLASROOLDWMEADKVAAGPLRSALPAGW-----FIAD------KSG-AGE- 214
DB 115 -----EMVR-----DLHGMVTVPDLTFLEFDEPIAVKRCGRGQTFK 154
QY 215 -----RGSRCITIALGPDKSPRIIVYITTGSOATMDE 247
DB 155 EKIEFLRGVRELEFLGLAE-EPERFVIADAGSPDEVER 192

RESULT 8
VNSC_RINDK
ID VNSC_RINDK STANDARD: PRT; 177 AA.
AC P35948;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein C.
GN C.
OS Rinderpest virus (strain Kabete O) (RDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID=11242;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92410646; PubMed=1529555;

RA Yamanaka M., Dale B., Crisp T., Cordell B., Grubman M., Ylma T.;
 RT "Sequence analysis and editing of the phosphoprotein (P) gene of
 RL rinderpest virus.";
 CC Virology 190:553-556(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S44819; AAB23269.1; -
 DR PIR: B43387; B43387.
 DR InterPro: IPR003875; Paramyxovir_NS_C.
 DR Pfam: PF02725; Paramyxo_NS_C; 1.
 KW Nonstructural protein.
 SO SEQUENCE 177 AA; 19926 MW; 116971B140A39F11 CRC64;

Query Match 4.8%; Score 65; DB 1; Length 177;
 Best Local Similarity 24.6%; Pred. No. 2.2e+02;
 Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;

OY 93 TVRELCSAATMSDNTANLLTTIGCPKELTAFLHNMGDHVTLDREPELNEAIPNDE 152
 DB 49 TIRISASHASQQLDQAKACLAVTIDLEATVMSR-----WEHSL----- 90
 OY 153 RDTTP-----VANATTLKLLTGELTLASRQGLIDWNEADKVAGPLLSALPA 202
 DB 91 --VTPQCIAPRVSTIMEFMTAVKRLRESKMLTSLSPNQL--MMVSR-SGEEMKMLRTA 144
 OY 203 GWFAD 208
 DB 145 MWILAN 150

RESULT 9
 ID VNSC_RINDR STANDARD; PRT; 177 AA.
 AC 003339;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Nonstructural protein C.
 GN C.
 OS Rinderpest virus (strain RBOC) (RDV).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
 OX NCBI_TaxID=36409;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9315558; PubMed=8429304;
 RA Baron M.D., Shaila M.S., Barrett T.;
 RT "Cloning and sequence analysis of the phosphoprotein gene of
 RT rinderpest virus.";
 RL J. Gen. Virol. 74:299-304(1993).
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68311; CA48391.1; -
 DR EMBL: Z30697; CA483179.1; -
 DR PIR: J01931; J01931.
 DR InterPro: IPR003875; Paramyxovir_NS_C.
 DR Pfam: PF02725; Paramyxo_NS_C; 1.
 KW Nonstructural protein.
 SO SEQUENCE 177 AA; 19926 MW; 76D8D46E46D3FB07 CRC64;

Query Match 4.8%; Score 65; DB 1; Length 177;
 Best Local Similarity 24.6%; Pred. No. 2.2e+02;
 Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;

OY 93 TVRELCSAATMSDNTANLLTTIGCPKELTAFLHNMGDHVTLDREPELNEAIPNDE 152
 DB 49 TIRISASHASQQLDQAKACLAVTIDLEATVMSR-----WEHSL----- 90
 OY 153 RDTTP-----VANATTLKLLTGELTLASRQGLIDWNEADKVAGPLLSALPA 202
 DB 91 --VTPQCIAPRVSTIMEFMTAVKRLRESKMLTSLSPNQL--MMVSR-SGEEMKMLRTA 144
 OY 203 GWFAD 208
 DB 145 MWILAN 150

RESULT 10

ID MARR_ECOLI STANDARD; PRT; 144 AA.
 AC P27245; P76882; P77582;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Multiple antibiotic resistance protein marr.
 GN MARR OR SOXQ OR CFXB OR INAR OR B1530.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93186717; PubMed=8383113;
 RA Cohen S.P., Haechler H., Levy S.B.;
 RT "Genetic and functional analysis of the multiple antibiotic
 RT resistance (marr) locus in Escherichia coli.";
 RL J. Bacteriol. 175:1484-1492(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426517; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1234-1238(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Albe H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [4]
 RP CHARACTERIZATION
 RX MEDLINE=94110219; PubMed=8282690;
 RA Atiza R.R., Cohen S.P., Bachhawat N., Levy S.B., Dimple B.;
 RT "Repressor mutations in the marrAB operon that activate oxidative
 RT stress genes and multiple antibiotic resistance in Escherichia
 RT coli.";
 RL J. Bacteriol. 176:143-148(1994).
 RN [5]
 RP CHARACTERIZATION
 RX MEDLINE=95286534; PubMed=7768850;
 RA Seoane A.S., Levy S.B.;

```

RF "Characterization of MarR, the repressor of the multiple antibiotic
RT resistance (mar) operon in Escherichia coli."
RL J. Bacteriol. 177:3414-3419(1995).
RN [6]
RP MUTAGENESIS.
RX MEDLINE-20223625; PubMed-10760140;
RA Alkshun M.N., Kim Y.S., Levy S.B.;
RT "Mutational analysis of MarR, the negative regulator of marAB
RT expression in Escherichia coli, suggests the presence of two regions
RT required for DNA binding."
RL Mol. Microbiol. 35:1394-1404(2000).
CC -1- FUNCTION: REPRESSOR OF THE MARAB OPERON WHICH IS INVOLVED IN THE
CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS
CC GENES. BINDS TO THE MARO OPERATOR/PROMOTER SITE.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96235; AAC16394.1; ALT_INIT.
DR EMBL; AE000250; AAC74603.1; ALT_INIT.
DR EMBL; D90795; BAA15212.1; -.
DR EMBL; D90796; BAA15220.1; -.
DR EMBL; D90797; BAA15232.1; -.
DR PIR; E64907; E64907.
DR PDB; 1JGS; 28-DEC-01.
DR Ecocore; EG11435; marR.
DR InterPro: IPR000835; HTH_MarR.
DR Pfam: PF01047; MarR; 1.
DR PRINTS; PR00598; HTHMAR.
DR SMART; SM00347; HTH_MARR; 1.
DR PROSITE; PS01117; HTH_MARR_FAMILY; 1.
DR Transcription regulation: DNA-binding; Repressor;
DR Antibiotic resistance: Complete proteome; 3D-structure.
FT MUTAGEN 45 45 V->E; INCREASED TRANSCRIPTION OF THE
FT REGION II TRANSCRIPT.
FT MUTAGEN 77 77 R->L; INCREASED TRANSCRIPTION OF THE
FT REGION II TRANSCRIPT.
FT MUTAGEN 123 144 MISSING: INCREASED TRANSCRIPTION OF THE
FT REGION II TRANSCRIPT.
SQ SEQUENCE 144 AA; 16065 MW; BE7DE5549E24D1D3 CRC64;
Query Match 4.7%; Score 63.5; DB 1; Length 144;
Best local similarity 29.5%; Pred. NO. 2.2e+02;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;
QY 66 LGRRIHY--SOND--LVEY--SP--VTEKHLDGMYVELCSAIIYMSDNTANLLTTIG 118
DB 14 LGRILVHVNOKKDKDLNEVLSPLDITAAQFKVLCSIR--CAACIT----- 56
QY 119 GPEELTFELH-NMGDHTRLDR-----WEPELNEAIPIRDERD-----TTTPVAMATTLRK 167
DB 57 -PEVLKVLVSVDLALTRMLDRLVCKGWERLPN--PNOKRGVLKLTGGAAICEDQHQ 113
QY 168 LITGCELTLASROOLDIMNEADKVA--GPLRSALP 201
DB 114 LVGODL-----HDELTKRLTADAEVATLEVLKKVLP 144

```

```

GN YCJC OR B1299 OR SF1304.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_Taxid=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES-E.coli; STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blittner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES-E.coli; STRAIN-K12;
RX MEDLINE-9725157; PubMed-9097039;
RA Aliba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miyi T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horinouchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE OF 65-185 FROM N.A.
RC SPECIES-E.coli;
RX MEDLINE-91216440; PubMed-1840553;
RA Helm R., Strehler E.E.;
RT "Cloning an Escherichia coli gene encoding a protein remarkably
RT similar to mammalian aldehyde dehydrogenases."
RL Gene 99:15-23(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES-S.flexneri; STRAIN-301 / Serotype 2a;
RX MEDLINE-22272406; PubMed-12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Sun L., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [5]
RP IDENTIFICATION.
RC SPECIES-E.coli;
RA Rudd K.E.;
RT Unpublished observations (Aug-1994).
RL -1- SIMILARITY: SOME, TO H.INFLUENZAE H10659.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000228; AAC74381.1; -.
DR EMBL; D90768; BAA14868.1; -.
DR EMBL; D90767; BAA14859.1; -.
DR EMBL; M38433; -. NOT ANNOTATED_CDS.
DR EMBL; AE015157; AAN42915.1; -.
DR PIR; E64878; E64878.
DR Ecocore; EG12431; ycjC.
DR InterPro: IPR007113; Cupin_sup.
DR InterPro: IPR001387; HTH_3.

```



```

RX MEDLINE-99287316; PubMed-10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson M.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (by similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B, HSLV SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001728; AAD35606.1; -.
DR PIR: G72365; G72365.
DR HSSP: P31059; 1E94.
DR MEROPS: T01.006; -.
DR TIGR: TW0521; -.
DR HAMAP: MF_00248; -. 1.
DR InterPro: IPR001353; Protase-Protease.
DR Pfam: PF00227; proteasome; 1.
DR Hydrobase: Protease; Complete proteome.
KW ACT_SITE 6
FT ACCT_SITE 6 BY SIMILARITY.
SQ SEQUENCE 176 AA; 18933 MW; ECCE369602A0ABD02 CRC64;

Query Match 4.7%: Score 63; DB 1; Length 176;
Best Local Similarity 28.2%: Pred. No. 3.2e+02;
Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

OY 117 IGGPKELTAFILHNMGDHVTLRDMEPELNEAIPNDERPTTPVAMATLRKLLTGELLTL 176
DB 42 IGGKVLAFAGSVADMTLFDREKLR-----WGNNLTK 78
OY 177 ASHQQLIDWMEADKVAQPLRSALPAGFIADK-----SGAGSGSRGII-----AAL 224
DB 79 AAVELAKDW-RTDRV-----LR-RLKALLLVADKENIFIIISGGE-----VIGPDDDAAI 127
OY 225 GPDG 228
DB 128 GSGG 131

RESULT 14
HGB_RABIT
AC P02099; STANDARD; PRT; 146 AA.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hemoglobin gamma chain (Beta-3).
GN HGB.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-82053017; PubMed-6271761;
RA Hardison R.C.;
RT "The nucleotide sequence of rabbit embryonic globin gene beta 3.";
RL J. Biol. Chem. 256:11780-11786(1981).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-89178632; PubMed-2486295;
RA Margot J.B., Demers G.W., Hardison R.C.;
RT "Complete nucleotide sequence of the rabbit beta-like globin gene
RT cluster. Analysis of intergenic sequences and comparison with the
RT human beta-like globin gene cluster.";
RL J. Mol. Biol. 205:15-40(1989).
CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE
CC GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE
CC HUMAN GAMMA CHAIN GENE, WHICH CODES FOR A FETAL GLOBIN.
CC -1- SUBUNIT: Heterotrimer of two alpha chains and two gamma chains.
CC -1- TISSUE SPECIFICITY: Red blood cells.
CC -1- SIMILARITY: Belongs to the globin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M18818; AAA02984.1; -.
DR PIR: A02417; HBRB3.
DR HSSP: P02100; 1A9W.
DR InterPro: IPR002337; Beta_haem.
DR InterPro: IPR000971; Globin.
DR Pfam: PF00042; globin; 1.
DR PRINTS: PR00814; BETAHAEM.
DR PROSITE: PS01033; GLOBIN; 1.
KW Heme; Oxygen transport; Transport; Erythrocyte; Embryo.
FT INIT_MEN 0
SQ SEQUENCE 146 AA; 16093 MW; EB8D6C1C24DD2D82 CRC64;

Query Match 4.6%: Score 62.5; DB 1; Length 146;
Best Local Similarity 30.8%: Pred. No. 2.7e+02;
Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6;

OY 35 FRSEERFPMSFVKVLLGAVLSRIDAGQEQGLRIHVSQNDLVESPYTERKLTGDMTV 94
DB 3 FFAEKAALITSTK-----LVDEYDAGAEKLGK-----LVYV-FPTQFPDSFGNL 48
OY 95 RELCSAATMSDNTAANLLTTIGGPRELTAFLHNMGDHVTRLD 138
DB 49 SS--SSAINGNPKVKAH-----GKRVLTAF-----GVAVNVD 79

RESULT 15
PDUV_SALTY
ID PDUV_SALTY STANDARD; PRT; 150 AA.
AC O9XDM6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Propanediol utilization protein pduv.
GN PDUV OR STM2056.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxId=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-LT2;
RX MEDLINE-99429843; PubMed-10498708;
RA Bobik T.A., Havemann G.D., Busch R.J., Williams D.S., Aldrich H.C.;
RT "The propanediol utilization (pdu) operon of Salmonella enterica
RT serovar typhimurium LT2 includes genes necessary for formation of
RT polyphedral degradation.";
RL J. Bacteriol. 181:5967-5975(1999).
RN [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE-21534948; PubMed-11677609;
 RA McCelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dame M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT *Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT2.
 RT Nature 413:852-856(2001).
 RL -1- INDUCTION: By propanediol.
 CC -1- SIMILARITY: BELONGS TO THE EUPV/PDVV FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF026270; AAD39020.1; -;
 DR EMBL: AE008790; AAL20960.1; -;
 DR StGene: SGI0680; pdyv.
 KW ATP-binding; Complete proteome.
 FT NP_BIND 8 15 ATP (POTENTIAL).
 SQ SEQUENCE 150 AA: 16348 MW: 7771229432F97E56 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 150;
 Best Local Similarity 19.2%; Pred. No. 2.8e+02;
 Matches 29; Conservative 20; Mismatches 75; Indels 27; Gaps 4;

OY 52 CGAVLSRIDAGQEQIGRIHYSQNDLVEYSPYTEKHLTDGMTVRELCSAAITMSDNTAAN 111
 DB 12 CG---RTSLTQSLRGEALHYKRTQAIEWSPMAID--TGEYLENRCIYSALTSAACEAD 65
 OY 112 LLTTTGGPKRELTAFLLHNGDHVTRLDREPELNEAIPNDERDTTPVAAATTLLKLTG 171
 DB 66 VIALVYNADAAQWSPF-----SPGFTAPNNRPTTIGLVTKADLAEPQRIISLVA 111
 OY 172 ELLTLASRQQLDWMWEMADKVAGPLLSALPA 202
 DB 112 EMLTQAGAGQIF-----ITSALNNSGIDA 135

Search completed: September 10, 2003, 12:20:51
 Job time : 6.57143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 25.4286 Seconds

(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_37S_157T
Perfect score: 1345
Sequence: 1 HPETLVKVKADQKGLKARVY.....TMDERNQIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 524 | 39.0 | 105 | 2 | 052026 pseudomonas |
| 2 | 524 | 39.0 | 105 | 2 | 052330 escherichia |
| 3 | 499 | 37.1 | 145 | 2 | 08RTD8 Klebsiella |
| 4 | 488 | 36.3 | 102 | 2 | 052639 pseudomonas |
| 5 | 464 | 34.5 | 138 | 2 | 0990F3 escherichia |
| 6 | 464 | 34.5 | 138 | 2 | 09AMA1 escherichia |
| 7 | 464 | 34.5 | 139 | 2 | 09AMA2 escherichia |
| 8 | 462 | 34.3 | 138 | 2 | 09AMA0 escherichia |
| 9 | 462 | 34.3 | 138 | 2 | 09AMA9 escherichia |
| 10 | 459 | 34.1 | 139 | 2 | 09AMA3 escherichia |
| 11 | 459 | 34.1 | 139 | 2 | 09AMA3 escherichia |
| 12 | 338 | 25.1 | 67 | 2 | 053553 shigella fl |
| 13 | 326 | 24.2 | 95 | 2 | 08GDE5 Klebsiella |
| 14 | 310 | 23.0 | 62 | 2 | 09JN58 shigella fl |
| 15 | 301 | 22.4 | 128 | 2 | 08VQ00 escherichia |
| 16 | 251 | 18.7 | 180 | 2 | 08KVT2 staphylococ |

| | | | | | | |
|----|-------|------|-----|----|--------|---------------------|
| 17 | 242 | 18.0 | 52 | 2 | 09RA12 | 09c412 shigella fl |
| 18 | 241 | 17.9 | 48 | 2 | 09RLH0 | 09c1h0 proteus mir |
| 19 | 197 | 14.6 | 38 | 2 | P97145 | P97145 escherichia |
| 20 | 173.5 | 12.9 | 134 | 2 | 08VUL3 | 08vul3 staphylococ |
| 21 | 128.5 | 9.6 | 109 | 2 | 053698 | 053698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | P97146 | P97146 escherichia |
| 23 | 87 | 6.5 | 100 | 2 | 093S05 | 093s05 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | 09XB2 | 09xb2 bacillus ce |
| 25 | 80.5 | 6.0 | 202 | 16 | 09B520 | 09b520 rhizobium l |
| 26 | 80 | 5.9 | 198 | 2 | 09ACM8 | 09acm8 streptococ |
| 27 | 78 | 5.6 | 181 | 16 | 08D108 | 08d108 yersinia in |
| 28 | 75 | 5.5 | 153 | 5 | 08S520 | 08s520 yersinia pe |
| 29 | 74.5 | 5.5 | 113 | 2 | 049970 | 049970 mycobacteri |
| 30 | 74.5 | 5.5 | 172 | 16 | 08TBW3 | 08tbw3 bruceella me |
| 31 | 74.5 | 5.5 | 172 | 16 | 08FWF9 | 08fwf9 yersinia en |
| 32 | 73 | 5.4 | 175 | 17 | 08X9H0 | 08x9h0 yersinia en |
| 33 | 73 | 5.4 | 175 | 17 | 08X9H0 | 08x9h0 yersinia en |
| 34 | 73 | 5.4 | 201 | 17 | 08TM84 | 08tm84 methanosarc |
| 35 | 72.5 | 5.4 | 145 | 2 | 005384 | 005384 staphylococ |
| 36 | 72.5 | 5.4 | 184 | 5 | 045379 | 045379 caenorhabdl |
| 37 | 72.5 | 5.4 | 200 | 16 | 097JER | 097jeh clostridium |
| 38 | 72 | 5.4 | 205 | 6 | 09N275 | 09n275 ovis aries |
| 39 | 72 | 5.4 | 205 | 16 | 09B118 | 09b118 rhizobium l |
| 40 | 71.5 | 5.3 | 152 | 16 | 09R191 | 09r191 streptomyce |
| 41 | 71.5 | 5.3 | 194 | 2 | 08G6G0 | 08g6g0 borrelia af |
| 42 | 71.5 | 5.3 | 195 | 16 | 09A1A0 | 09a1a0 streptomyce |
| 43 | 71.5 | 5.3 | 196 | 2 | 09A1A0 | 09a1a0 streptomyce |
| 44 | 71 | 5.3 | 150 | 10 | 09A1G4 | 09a1g4 nictotiana t |
| 45 | 71 | 5.3 | 153 | 16 | 09PKT4 | 09pkt4 chlamydia m |

ALIGNMENTS

RESULT 1

052026 PRELIMINARY; PRT; 105 AA.
ID 052026
AC 052026;
DT 01-NOV-1996 (TRENBLER, 01, Created)
DT 01-NOV-1996 (TRENBLER, 01, Last sequence update)
DT 01-MAR-2002 (TRENBLER, 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN BLA.
OS Pseudomonas aeruginosa.
OC Plasmid PRO1614.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95011664; PubMed=7926843;
RA West S.E., Schweitzer H.P., Dail C., Sample A.K., Runyen-Janecky L.J.;
RT "Construction of Improved Escherichia-Pseudomonas shuttle vectors
RT derived from pUC18/19 and sequence of the region required for their
RT replication in Pseudomonas aeruginosa.";
RL Gene 148:81-86(1994).
DR EMBL: L30112; AAA66058.1;
DR HSSP; P0810; IXPB.
DR InterPro; IPR001466; Beta_lactamase.
DR InterPro; IPR000871; Beta_lactamase_A.
DR Pfam; PF00144; beta_lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
KW plasmid.
FT NON_TER
SQ SEQUENCE 105 AA; 11229 MW; D2889AA07330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.2e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTIRKLTGCELTTLASROOLIDWNEADKVGPIRLSALPACWFTADKSGAGRGSRG 219
DB 2 AMATTIRKLTGCELTTLASROOLIDWNEADKVGPIRLSALPACWFTADKSGAGRGSRG 61

QY 220 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 263
DB 62 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 105

RESULT 2
O52330 PRELIMINARY; PRT: 105 AA.

AC 052330; 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Th1 bla protein (Fragment).
OS Escherichia coli.
OG Plasmid RK2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;

RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-90264294; Pubmed-2160936;
RX Kornacki J.A., Burlage R.S., Figurski D.H.;
RT "The *kil-kor* regulon of broad host-range plasmid RK2: Nucleotide
RT sequence, polypeptide product and expression of regulatory gene
RT *korc*.";
RL J. Bacteriol. 172:3040-3050(1990).
DR EMBL: M32794; AAA26408.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_Lactamase.
DR InterPro: IPR000871; Beta_Lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
KW NON_TER 1 1
SQ SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.2e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMATLRKLLTGELITLASRQQLIDMMEADKVGAPLRSALPAGWFIADKSGAGERSRG 219
DB 2 AAMATLRKLLTGELITLASRQQLIDMMEADKVGAPLRSALPAGWFIADKSGAGERSRG 61

QY 220 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 263
DB 62 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 105

RESULT 3
O8RTD8 PRELIMINARY; PRT: 145 AA.

AC 08RTD8; 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE SHV-5 enzyme (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;

RN [1]
RP SEQUENCE FROM N.A.
RA Palaeubramaniam S.;
RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
RT cefazidime-resistant Klebsiella pneumoniae.";
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF467105; AAV5506.1; -.
DR InterPro: IPR001466; Beta_Lactamase.
DR Pfam: PF00144; beta-lactamase; 1.
DR NON_TER 1 1
FT NON_TER 145 145

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 37.1%; Score 499; DB 2; Length 145;
Best Local Similarity 68.8%; Pred. No. 4.8e-35;
Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 100 AATMSDNFTANLLTTIGGKPELTAFLNNQDHYRLDRWPELNEAIPNDRDFTTPV 159
DB 7 AATMSDNFTANLLTAIVGGAGLTAFLRQIGDNTYRLDRWPELNEALPGDARDTTTPA 66

QY 160 AAMATLRKLLTGELITLASRQQLIDMMEADKVGAPLRSALPAGWFIADKSGAGERSRG 219
DB 67 AAMATLRKLLTGELITLASRQQLIDMMEADKVGAPLRSALPAGWFIADKSGAGERSRG 126

QY 220 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 263
DB 127 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 144

RESULT 4
O52639 PRELIMINARY; PRT: 102 AA.

AC 052639; 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Beta-Lactamase (Fragment).
GN BLA.
OS Pseudomonas aeruginosa.
OC Plasmid PRO1600.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;

RN [1]
RP SEQUENCE FROM N.A.
RA TRANSPOSON-Tn1.
RX MEDLINE-94336757; Pubmed-8058819;
RX Jansons I., Touchie G., Sharp R., Almqvist K., Farinha M.A., Lam J.S.,
RT "Deletion and transposon mutagenesis and sequence analysis of the *pQR*
RT PRO1600 *Orr* region found in the broad-host-range plasmids of the *pQR*
RT series.";
RL Plasmid 31:265-274(1994).
DR EMBL: L22691; AAA98312.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_Lactamase.
DR InterPro: IPR000871; Beta_Lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
KW NON_TER 1 1
SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.3%; Score 488; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 2.5e-34;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAMATLRKLLTGELITLASRQQLIDMMEADKVGAPLRSALPAGWFIADKSGAGERSRG 219
DB 2 AAMATLRKLLTGELITLASRQQLIDMMEADKVGAPLRSALPAGWFIADKSGAGERSRG 61

QY 220 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 263
DB 62 IIAALGPDGKPSRIVYITGSGQATMDERNRQIAETGASLIRKM 105

RESULT 5
O990F3 PRELIMINARY; PRT: 138 AA.

AC 0990F3; 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)

RC STRAIN-E/98 9-1;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327052; AAK07467.1; -.
DR HSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 6.7e-32;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLNMGDHYTRLDRWEPELNPAIPNDERDTTPVAMATTIRKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTPASMAATTIRKLLT 69
OY 171 GELLTLASRQOLIDWMEADKVAAPLRSALPAGWFIADKSGAGSGRSGIITAAAGPDGKP 230
DB 70 SORLSARSQROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGANGIYALLGPNNKA 129
OY 231 SRIVVY 237
DB 130 ERIVLV 136

RESULT 9
O9AM99 PRELIMINARY; PRT; 138 AA.

ID O9AM99; PRELIMINARY; PRT; 138 AA.
AC 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/99 4-1;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327053; AAK07468.1; -.
DR HSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 6.7e-32;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLNMGDHYTRLDRWEPELNPAIPNDERDTTPVAMATTIRKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTPASMAATTIRKLLT 69
OY 171 GELLTLASRQOLIDWMEADKVAAPLRSALPAGWFIADKSGAGSGRSGIITAAAGPDGKP 230
DB 70 SORLSARSQROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGANGIYALLGPNNKA 129

OY 231 SRIVVY 237
DB 130 ERIVLV 136

RESULT 10

ID O9AM98 PRELIMINARY; PRT; 139 AA.
AC O9AM98;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/99 3-2;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327054; AAK07469.1; -.
DR HSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 139 AA; 15234 MW; 0361A792F5442A8 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 6.8e-32;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLNMGDHYTRLDRWEPELNPAIPNDERDTTPVAMATTIRKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTPASMAATTIRKLLT 69
OY 171 GELLTLASRQOLIDWMEADKVAAPLRSALPAGWFIADKSGAGSGRSGIITAAAGPDGKP 230
DB 70 SORLSARSQROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGANGIYALLGPNNKA 129
OY 231 SRIVVY 237
DB 130 ERIVLV 136

RESULT 11

ID O9AMA3 PRELIMINARY; PRT; 139 AA.
AC O9AMA3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-935;
RA Subramaniam G., Navaratnam P.;
RT SHV gene sequences from cefotaxidime-resistant *Escherichia coli*
RT Isolates from UMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327047; AAK07462.1; -.
DR HSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.

OY 111 NLLTTTGGPKELTAFLNMGDHYTRLDRWEPELNPAIPNDERDTTPVAMATTIRKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTPASMAATTIRKLLT 69
OY 171 GELLTLASRQOLIDWMEADKVAAPLRSALPAGWFIADKSGAGSGRSGIITAAAGPDGKP 230
DB 70 SORLSARSQROLQWVDDRVAGPLIRSVLPAGWFIADKSGKRGANGIYALLGPNNKA 129

```

DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Query Match
Best Local Similarity 67.7%; Score 459; DB 2; Length 139;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NULLTTGGPKELTATLHNGDHYRLDRPEPELNEAIPNDERDTTPVAMAATTLKLLT 170
DB 10 HLLATVGGPAGTATLROIGDVTTLDRWETELNEALPGARDTTPASMAATTLKLLT 69
QY 171 GELLTLASRQOLIDMNEADVPILRSALPGWFIADKSGGERSGCIIALGPDGR 230
DB 70 SQRLSARSQROLIDMNEADVPILRSALPGWFIADKSGGERSGCIIALGPDGR 129
QY 231 SRIVVY 237
DB 130 ERIVVY 136

RESULT 12
ID 053553 PRELIMINARY; PRT; 67 AA.
AC 053553.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE-96081517; PubMed-8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
EMBL: S81098; AAB35839.2; .
DR HSSP; P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 67
SQ SEQUENCE 67 AA; 7287 MW; 85F8C5B0363FCB CRC64;

Query Match
Best Local Similarity 25.1%; Score 338; DB 2; Length 67;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 55 VLSRDAGOEQGRRIHYSQNDLYEYSPTEKHITDGMVVELCSAATMSDNTANLL 114
DB 1 VLSRDAGOEQGRRIHYSQNDLYEYSPTEKHITDGMVVELCSAATMSDNTANLL 60
QY 115 TTIGPK 121
DB 61 TTIGPK 67

RESULT 13
ID 08GDE5 PRELIMINARY; PRT; 95 AA.
AC 08GDE5;

```

```

DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Bellio H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Amves S.G.B.;
RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
RT Chile."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY150585; AAN7730.1; .
FT NON_TER 1
FT NON_TER 95
SQ SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
Best Local Similarity 24.2%; Score 326; DB 2; Length 95;
Matches 62; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 147 AIPNDERDTTPVAMAATTLKLLTGELITLASRQOLIDMNEADVPILRSALPGWFI 206
DB 1 ALPGDARDTTPASMAATTLKLLTSQRLSARSQROLIDMNEADVPILRSALPGWFI 60
QY 207 ADKSGAGERSGCIIALGPDGKPSRIYVITGS 241
DB 61 ADKTSERGARGIVALLGPNNAERYIVYLRDS 95

RESULT 14
ID 09JN58 PRELIMINARY; PRT; 62 AA.
AC 09JN58.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE-96081517; PubMed-8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
EMBL: S81099; AAB35840.1; .
DR HSSP; P00810; 1XPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1
FT NON_TER 62
SQ SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
Best Local Similarity 23.0%; Score 310; DB 2; Length 62;
Matches 60; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 132 DHVTRLDREPELNEAIPNDERDTTPVAMAATTLKLLTGELITLASRQOLIDMNEADVP 191
DB 1 DHVTRLDREPELNEAIPNDERDTTPVAMAATTLKLLTGELITLASRQOLIDMNEADVP 60
QY 192 AG 193

```

Db 61 AG 62

RESULT 15

Q8V000

PRELIMINARY; PRT; 128 AA.

AC Q8V000:

DT 01-MAR-2002 (TREMblrel. 20, Created)

DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)

DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)

DE Ctx-M type beta-lactamase (Fragment).

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN-clinical isolate HD3;

RA Stuerenburg E., Feucht H., Laufs R.;

RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF466148; AAL73983.1; -

DR InterPro: IPR001466; Beta_lactamase.

DR InterPro: IPR000871; Beta_lactamase_A.

DR Pfam: PF00144; beta-lactamase; 1.

DR PRINTS: PRO018; BLACTAMASEA.

FT NON_TER 1

FT NON_TER 128

SQ SEQUENCE 128 AA; 13878 MW; F2EB815B84FF318A CRC64;

Query Match

22.4%; Score 301; DB 2; Length 128;

Best local similarity 48.4%; Pred. No. 3.8e-18;

Matches 59; Conservative 20; Mismatches 43; Indels 0; Gaps 0;

QY 66 LGRRIHYSQNDLVEYSPVTEKHLTDGNTVRELCSAATITMSDNTAANLLTTIGCPKELTA 125

DB 6 LNRVETIKKSDLVNYPNPIAEKHVNGTMSLAELISAAALQSDNVAMNKLIAHVGGPASVTA 65

QY 126 FLNNMGDHYVRLDRWEPELNEAIPNDERDTTPVAMATTLRKLLTGELLTLASROQLIDW 185

DB 66 FARQLDDETRLDRTETTLNLTALPQDPDRTISPRAMAQTLRNLTIGKALGDSQRAQLVTH 125

QY 186 ME 187

DB 126 MK 127

QY 186 ME 187

DB 126 MK 127

Search completed: September 10, 2003, 12:29:28
Job time : 25.4286 secs

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251969.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-483426/52.
XX
DR N-PSDB: AAK63652.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Claim 11; SEQ ID NO 18464; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 129 AA;
Query Match 34.9%; Score 469; DB 22; Length 129;
Best Local Similarity 95.8%; Pred. No. 1.4e-40;
Matches 92; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 HPEILVVKDAEDDQAGRVGYIELDNGSEILSFSESEFPFPMSTFKVLLCGAVLSRID 60
DB 28 HPEILVVKDAEDDQAGRVGYIELDNGSEILSFSESEFPFPMSTFKVLLCGAVLSRID 87
QY 61 AGGQQLGRRIRHSQNDLVEYSPTVEKHLTDGMYRE 96
DB 88 AGGQQLGRRIRHSQNDLVEYSPTVEKHLTDGMYRE 123
RESULT 7
ID AAG27917 standard; Protein; 182 AA.
XX
AC AAG27917;

XX
XX
DT 18-FEB-2002 (first entry)
XX
XX
DE Novel human diagnostic protein #27908.
XX
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensics;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB: AAS92104.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 20; SEQ ID NO 58276; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-AAG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 182 AA;
Query Match 34.9%; Score 469; DB 22; Length 182;
Best Local Similarity 86.4%; Pred. No. 2.3e-40;
Matches 95; Conservative 4; Mismatches 7; Indels 4; Gaps 1;
QY 1 HPEILVVKDAEDDQAGRVGYIELDNGSEILSFSESEFPFPMSTFKVLLCGAVLSRID 60
DB 24 HPEILVVKDAEDDQAGRVGYIELDNGSEILSFSESEFPFPMSTFKVLLCGAVLSRID 83
QY 61 AGGQQLGRRIRHSQNDLVEYSPTVEKHLTDGMYRE 110
DB 84 AGGQQLGRRIRHSQNDLVEYSPTVEKHLTDGMYRE 129
RESULT 8
ID AAB59052 standard; Protein; 94 AA.
XX
AC AAB59052

PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 05-SEP-2000; 2000US-0230437.
 PR 05-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233066.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HMDA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-465566/50.
 DR N-PSDB; AAS41090.
 DR
 DR
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 11; SEQ ID NO 1216; 1180bp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 129 AA;
 Query Match 34.9%; Score 469; DB 22; Length 129;
 Best Local Similarity 95.8%; Pred. No. 1.4e-40;
 Matches 92; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246530.
PR 08-NOV-2000; 2000US-0246531.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246534.
PR 08-NOV-2000; 2000US-0246535.
PR 08-NOV-2000; 2000US-0246536.
PR 08-NOV-2000; 2000US-0246537.
PR 08-NOV-2000; 2000US-0246538.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246540.
PR 08-NOV-2000; 2000US-0246541.
PR 08-NOV-2000; 2000US-0246542.
PR 08-NOV-2000; 2000US-0246543.
PR 08-NOV-2000; 2000US-0246544.
PR 08-NOV-2000; 2000US-0246545.
PR 08-NOV-2000; 2000US-0246546.
PR 08-NOV-2000; 2000US-0246547.
PR 08-NOV-2000; 2000US-0246548.
PR 08-NOV-2000; 2000US-0246549.
PR 08-NOV-2000; 2000US-0246550.
PR 08-NOV-2000; 2000US-0246551.
PR 08-NOV-2000; 2000US-0246552.
PR 08-NOV-2000; 2000US-0246553.
PR 08-NOV-2000; 2000US-0246554.
PR 08-NOV-2000; 2000US-0246555.
PR 08-NOV-2000; 2000US-0246556.
PR 08-NOV-2000; 2000US-0246557.
PR 08-NOV-2000; 2000US-0246558.
PR 08-NOV-2000; 2000US-0246559.
PR 08-NOV-2000; 2000US-0246560.
PR 08-NOV-2000; 2000US-0246561.
PR 08-NOV-2000; 2000US-0246562.
PR 08-NOV-2000; 2000US-0246563.
PR 08-NOV-2000; 2000US-0246564.
PR 08-NOV-2000; 2000US-0246565.
PR 08-NOV-2000; 2000US-0246566.
PR 08-NOV-2000; 2000US-0246567.
PR 08-NOV-2000; 2000US-0246568.
PR 08-NOV-2000; 2000US-0246569.
PR 08-NOV-2000; 2000US-0246570.
PR 08-NOV-2000; 2000US-0246571.
PR 08-NOV-2000; 2000US-0246572.
PR 08-NOV-2000; 2000US-0246573.
PR 08-NOV-2000; 2000US-0246574.
PR 08-NOV-2000; 2000US-0246575.
PR 08-NOV-2000; 2000US-0246576.
PR 08-NOV-2000; 2000US-0246577.
PR 08-NOV-2000; 2000US-0246578.
PR 08-NOV-2000; 2000US-0246579.
PR 08-NOV-2000; 2000US-0246580.
PR 08-NOV-2000; 2000US-0246581.
PR 08-NOV-2000; 2000US-0246582.
PR 08-NOV-2000; 2000US-0246583.
PR 08-NOV-2000; 2000US-0246584.
PR 08-NOV-2000; 2000US-0246585.
PR 08-NOV-2000; 2000US-0246586.
PR 08-NOV-2000; 2000US-0246587.
PR 08-NOV-2000; 2000US-0246588.
PR 08-NOV-2000; 2000US-0246589.
PR 08-NOV-2000; 2000US-0246590.
PR 08-NOV-2000; 2000US-0246591.
PR 08-NOV-2000; 2000US-0246592.
PR 08-NOV-2000; 2000US-0246593.
PR 08-NOV-2000; 2000US-0246594.
PR 08-NOV-2000; 2000US-0246595.
PR 08-NOV-2000; 2000US-0246596.
PR 08-NOV-2000; 2000US-0246597.
PR 08-NOV-2000; 2000US-0246598.
PR 08-NOV-2000; 2000US-0246599.
PR 08-NOV-2000; 2000US-0246600.

(HUMAN) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-483426/52.
N-PSDB: AAK63651.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

Claim 11: SEQ ID NO 18463; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM82191. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
suppress the activity of (I) by expressing inactive proteins or to
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/hematopoietic-related diseases, especially
cancers and cancer metastases of hematopoietic-derived cells. AAK64703
to AAK87654 represent human immune/hematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAM82169
represent sequences used in the exemplification of the present invention.

| | |
|----|--|
| CC | diagnosis, treatment, prevention and/or prognosis of a wide range of |
| CC | disorders including hyperproliferative disorders (e.g. cancer), |
| CC | immunodeficiency disorders (e.g. AIDS) autoimmune disorders |
| CC | (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), |
| CC | metabolic disorders (e.g. phenylketonuria), inflammatory disorders |
| CC | (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), |
| CC | blood-related disorders (e.g. haemophilia), reproductive disorders |
| CC | (e.g. infertility) and infectious disorders (e.g. influenza). The |
| CC | polynucleotides of the invention can also be used in gene therapy. |
| CC | AAU22915-AAU2314 represent the novel human enzyme polypeptides of the |
| CC | invention. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification, but was obtained in electronic format directly from WIPO |
| CC | at ftp.wipo.int/pub/published_pct_sequences. |
| CC | |
| XX | Sequence 146 AA: |
| XX | |
| XX | Query Match 41.9%; Score 563; DB 22; Length 146; |
| XX | Best Local Similarity 96.5%; Pred. No. 2,8e-50; |
| XX | Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0. |
| QY | 1 HPETLVKVKDAEDQLGARVGYIELDLNSGELLEFRSEERFPMKSTFKVLTCGAVLSRID 60 |
| DB | 28 HPETLVKVKDAEDQLGARVGYIELDLNSGILESPFEERFPMKSTFKVLTCGAVLSRID 87 |
| QY | 61 AGDEQLGRRIHYSQNDLVEYSPVTEKHLTDGTVRELGSAITMSDNTAANLLLT 115 |
| DB | 88 AGDEQLGRRIHYSQNDLVEYSPVTEKHLTDGTVRELGSAITMSDNTAANLLLT 142 |
| XX | RESULT 3 |
| XX | AA090870 |
| XX | AA090870 standard; Protein; 146 AA. |
| XX | AA090870: |
| DT | 07-NOV-2001 (first entry) |
| XX | |
| XX | Human Immune/haematopoietic antigen SEQ ID NO:18463. |
| XX | |
| XX | Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer; |
| XX | cytostatic; gene therapy; vaccine; metastasis. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | MO200157182-A2. |
| XX | |
| PD | 09-AUG-2001. |
| XX | |
| PF | 17-JAN-2001; 2001WO-US01354. |
| XX | |
| XX | 31-JAN-2000; 2000US-0179065. |
| PR | 04-FEB-2000; 2000US-0180628. |
| PR | 24-FEB-2000; 2000US-0184664. |
| PR | 02-MAR-2000; 2000US-0186350. |
| PR | 16-MAR-2000; 2000US-0189874. |
| PR | 17-MAR-2000; 2000US-0190076. |
| PR | 18-APR-2000; 2000US-0198123. |
| PR | 19-MAY-2000; 2000US-0205515. |
| PR | 07-JUN-2000; 2000US-0209467. |
| PR | 28-JUN-2000; 2000US-0214886. |
| PR | 30-JUN-2000; 2000US-0215135. |
| PR | 07-JUL-2000; 2000US-0216647. |
| PR | 07-JUL-2000; 2000US-0216880. |
| PR | 11-JUL-2000; 2000US-0217487. |
| PR | 11-JUL-2000; 2000US-0217496. |
| PR | 14-JUL-2000; 2000US-0218290. |
| PR | 26-JUL-2000; 2000US-0220963. |
| PR | 26-JUL-2000; 2000US-0220964. |
| PR | 14-AUG-2000; 2000US-0224518. |
| PR | 14-AUG-2000; 2000US-0224519. |
| PR | 14-AUG-2000; 2000US-0225213. |
| PR | 14-AUG-2000; 2000US-0225214. |

PA (EART) EARTH CHEM CO LTD.
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A,
PI Matsushiro S;
XX
DR WPI: 1986-015031/03.
XX N-PSDB; AAN60631.
XX
PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PS and transformed cells contg. it.
PS Disclosure: Page 56-59; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 196 AA:
SQ
Query Match 43.0%; Score 579; DB 7; Length 196;
Best Local Similarity 97.4%; Pred. No. 9, 3e-52;
Matches 114; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 HPETLVKKAEDQLGARVGTIELDLSNGEILSRSEFRFPMSDFKYLICAVLSRID 60
Db |||||
24 HPETLVKKAEDQLGARVGTIELDLSNGRIIESFRPEERFPMSDFKYLICAVLSRID 83
QY 61 AGGEOLGRIRIHSONDLVEGSPVTEKHITDGTFARELCSAATITMSDNTANILTTI 117
Db |||||
84 AGGEOLGRIRIHSONDLVEGSPVTEKHITDGTFARELCSAATITMSDNTANILTTI 140
RESULT 2
AAU23221
ID AAU23221 standard; Protein: 146 AA.
XX
XX AAU23221:
AC
XX 18-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #307.
DE
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
neurotropic; anticoagulant.
XX
XX Homo sapiens.
OS
XX
XX WO200155301-A2.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01239.
PE
XX
XX 31-JAN-2000; 2000US-0179065.
PR
XX 04-FEB-2000; 2000US-0180628.
PR
XX 24-FEB-2000; 2000US-0184664.
PR
XX 02-MAR-2000; 2000US-0186350.
PR
XX 16-MAR-2000; 2000US-0189874.
PR
XX 17-MAR-2000; 2000US-0190076.
PR
XX 18-APR-2000; 2000US-0198123.
PR
XX 19-MAY-2000; 2000US-0205515.
PR
XX 07-JUN-2000; 2000US-0209467.
PR
XX 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232196.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

Title: SE02_30E_37S_157T
Perfect score: 1345
Sequence: 1 HPETLVKVKAEQDLGARVG.....TMDERHROIAGISLIRKM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/gensseq/gensseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|---------------------|
| 1 | 579 | 43.0 | 196 | 7 | AAP60627 | Beta-urogastrone - |
| 2 | 563 | 41.9 | 146 | 22 | AAU33221 | Novel human enzyme |
| 3 | 563 | 41.9 | 146 | 22 | AAU33221 | Human immune/haema |
| 4 | 484 | 36.0 | 127 | 18 | AAW20440 | H. pylori cytoplasm |
| 5 | 469 | 34.9 | 129 | 22 | AAU33220 | Novel human enzyme |
| 6 | 469 | 34.9 | 129 | 22 | AAU33220 | Human immune/haema |
| 7 | 469 | 34.9 | 182 | 22 | ABG27917 | Novel human diapho |
| 8 | 457 | 34.0 | 94 | 21 | AAU59052 | Breast and ovarian |
| 9 | 396 | 29.4 | 159 | 7 | AAP60628 | Beta-urogastrone - |

| | | | | | | |
|----|------|------|-----|----|-----------|---------------------|
| 10 | 366 | 27.2 | 101 | 22 | ABG27915 | Novel human diapho |
| 11 | 325 | 24.2 | 88 | 22 | ABG27919 | Novel human diapho |
| 12 | 191 | 14.2 | 119 | 7 | AAP60628 | Beta-urogastrone - |
| 13 | 141 | 10.5 | 51 | 21 | AAU92783 | Vtgs-beta-lactama |
| 14 | 80.5 | 6.0 | 158 | 22 | AAU45138 | Protonibacterium |
| 15 | 79 | 5.9 | 200 | 18 | AAU55524 | H. pylori ORF 29p |
| 16 | 75 | 5.6 | 202 | 24 | AAU32763 | Human zcyto24 prot |
| 17 | 75 | 5.6 | 202 | 24 | AAU16283 | Mouse IMX129840-2 |
| 18 | 72.5 | 5.4 | 196 | 21 | AAU73344 | HTRM clone 0258181 |
| 19 | 71 | 5.3 | 159 | 19 | AAU69170 | N-terminally tagged |
| 20 | 70.5 | 5.2 | 193 | 23 | AAU76854 | Human integrin alp |
| 21 | 70.5 | 5.2 | 193 | 23 | AAU76853 | Human integrin alp |
| 22 | 70 | 5.2 | 15 | 10 | AAU98503 | Sequence encoded b |
| 23 | 70 | 5.2 | 170 | 22 | AAU52472 | Protonibacterium |
| 24 | 70 | 5.2 | 202 | 22 | AAU62686 | S. epidermidis ope |
| 25 | 69.5 | 5.2 | 202 | 22 | AAU66330 | Protonibacterium |
| 26 | 69.5 | 5.2 | 184 | 21 | AAU35446 | Plus radiata cell |
| 27 | 69 | 5.1 | 134 | 21 | AAU37220 | Zee mays protein f |
| 28 | 69 | 5.1 | 166 | 23 | ABP26537 | Streptococcus poly |
| 29 | 69 | 5.1 | 184 | 22 | AAU91764 | S. epidermidis ope |
| 30 | 68 | 5.1 | 180 | 21 | AAU20206 | Arabidopsis thalia |
| 31 | 67.5 | 5.0 | 152 | 22 | ABG19138 | Novel human diapho |
| 32 | 67.5 | 5.0 | 152 | 22 | ABG27771 | S. pneumoniae SGHR |
| 33 | 67.5 | 5.0 | 170 | 19 | AAU38722 | Human ORF1346 |
| 34 | 67.5 | 5.0 | 177 | 21 | AAU841582 | Human MK61 protein |
| 35 | 67.5 | 5.0 | 187 | 23 | AAU97104 | C glutamicum prote |
| 36 | 67 | 5.0 | 178 | 22 | AAU91926 | Protonibacterium |
| 37 | 66.5 | 4.9 | 83 | 22 | AAU59092 | Dermatophagoides p |
| 38 | 66 | 4.9 | 181 | 20 | AAU50544 | Synthetic house du |
| 39 | 66 | 4.9 | 181 | 22 | AAU19147 | S. pneumoniae derl |
| 40 | 66 | 4.9 | 181 | 22 | AAU19147 | Human zcyto25 prot |
| 41 | 66 | 4.9 | 183 | 19 | AAU85905 | Human alpha 2 C4 a |
| 42 | 66 | 4.9 | 202 | 24 | AAU32764 | Human dehydrogenas |
| 43 | 65.5 | 4.9 | 102 | 23 | ABP34479 | Partial corn extra |
| 44 | 65.5 | 4.9 | 102 | 23 | ABP34479 | |
| 45 | 65.5 | 4.9 | 136 | 21 | AAU44229 | |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AAP60627 | standard; Protein; 196 AA. |
| ID | AAP60627 |
| XX | |
| AC | AAP60627 |
| XX | |
| DT | 25-MAR-2003 (updated) |
| DT | 17-JUN-1991 (first entry) |
| XX | |
| DE | Beta-urogastrone - beta-lactamase fusion protein from pUC2101. |
| XX | |
| KW | Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds; |
| KW | fusion protein; beta-lactamase. |
| XX | |
| OS | Synthetic. |
| XX | |
| FT | Key |
| FT | Protein |
| FT | Misc-difference |
| FT | Protein |
| FT | Protein |
| FT | Protein |
| XX | |
| PN | DE3523634-A. |
| XX | |
| PD | 09-JAN-1986. |
| XX | |
| PF | 02-JUL-1985; 85DE-3523634. |
| XX | |
| PR | 02-JUL-1984; 84UP-0137691. |
| XX | |

Novel human diapho
Novel human diapho
Beta-urogastrone -
Vtgs-beta-lactama
Protonibacterium
H. pylori ORF 29p
Human zcyto24 prot
Mouse IMX129840-2
HTRM clone 0258181
N-terminally tagged
Human integrin alp
Human integrin alp
Sequence encoded b
Protonibacterium
S. epidermidis ope
Protonibacterium
Plus radiata cell
Zee mays protein f
Streptococcus poly
S. epidermidis ope
Arabidopsis thalia
Novel human diapho
S. pneumoniae SGHR
Human ORF1346
Human MK61 protein
C glutamicum prote
Protonibacterium
Dermatophagoides p
Synthetic house du
S. pneumoniae derl
Human zcyto25 prot
Human alpha 2 C4 a
Human dehydrogenas
Partial corn extra

Db 61 AG 62

RESULT 15

08VQ00

| | | | | |
|----|----------|--------------|------|---------|
| ID | Q8VQ000 | PRELIMINARY; | PRT; | 128 AA. |
| AC | Q8VQ000- | | | |

01-MAR-2002 (Tremblay, 20. Created by

DT 01-MAR-2002 (Tremblé, 20, Last sequence update)

01-MAR-2003 (TREMBLER, 23, Last annotation update)

05 Escherichia coli.

Bacteria; Proteobacteria

Enterobacteriaceae; Escherichia.
NCBI TaxID=562:

| | |
|----|-----|
| RN | [1] |
|----|-----|

| REP | SEQ |
|-----|-----|
| RC | STP |

RA Stuerenburg F...

^aA new CTX-M type beta-lactamase

Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
EMBL: AF456148.1; F073003.1
DR

InterPro; IPR001466: Beta lactamase

InterPro; IPR000871; Beta_lactamase A

PRINTS: PR00119, PR00120, PR00144; beta-lactamase; 1.

| FT | NON_TER | 1 | 1 |
|----|---------|---|---|
| FT | NON_TER | 1 | 1 |

| | | |
|---------|-----|-----|
| NON_TER | 128 | 128 |
| FT | 128 | 128 |

SEQUENCE 128 AA; 13878 MW; F2EB81G

| Query Match | 22.28% | Score |
|-------------|--------|-------|
|-------------|--------|-------|

Best Local Similarity 48.4%;
Matchbox F0.0; G0.0

maçãs 39; Conservação

66 LGRIHYSNDLY

6 INOVEFECOT UNWNOE

| | | | | |
|-----------------------|--------|------------|--------|-------------|
| Query Match | 22.28; | Score 299; | DB 2; | Length 128; |
| Best Local Similarity | 48.48; | Pred No 1 | 40-17; | |

| | | | | | | | |
|---------|-----|--------------|-----|------------|-----|--------|---|
| Matches | 59; | Conservative | 19; | Mismatches | 44; | Indels | 1 |
|---------|-----|--------------|-----|------------|-----|--------|---|

66 ICBBTUYCNDIYEDVOCIMNNUUZZZ

[illegible]

6 LNQRVEIKKSDLVYNPIAEKHVNGTMSLAELSAALOYSDNVAMNKLI AHVGGPASYTA 65

126 FLHNMGDHYTRI DBEERET AEA TOWEEEEE.....
-----OOOAAAO OJ

185 IDQLSRQTLTCELLTGELTKRLKLVAMATIMPRDITNDEIPNEAIFELNEMWYANODIT

66 FAROLGDETFRLDRTETTLNTAIPGDPDRTSPRMAOTLRNLTIGKALGDSORAOI.VTW 125

186 ME 187

126 MK 127

Search completed: September 10, 2003, 12:29:28
Job time : 26.4286 secs

Job time : 26.4286 secs

```
QY      132 DHVTRLDKRWPELNEAIPNDERDFTMPAAATTLRLLTGELLTLASROOLDIMMEADKY 191
        |||||
Db       1 DHTRTLDKRWPELNEAIPNDERDFTMPAAATTLRLLTGELLTLASROOLDIMMEADKY 60
QY      192 AG 193
        ||
```


OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 263
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 105

RESULT 2

OY 052330 PRELIMINARY; PRT; 105 AA.
 AC 052330;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, last annotation update)
 DE Tbl bia protein (Fragment).
 OS Escherichia coli.
 OC plasmid RK2.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90264294; PubMed=2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT "The kil kor regulon of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT korC.";
 RL J. Bacteriol. 172:3040-3050(1990).
 DR EMBL: M32794; AAA26408.1; -
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase_1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR Plasmid.
 KW NON_TER

SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;
 SQ

Query Match 39.0%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1.1e-36;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AAMTTRKLLTGLTGLTASRQQLIDMNEADRVAGPLRSALPAGWFIADKSGAGERSRG 219
 DB 2 AAMTTRKLLTGLTGLTASRQQLIDMNEADRVAGPLRSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 263
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 105

RESULT 3

OY 08RTD8 PRELIMINARY; PRT; 145 AA.
 AC 08RTD8;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, last annotation update)
 DE SHV-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Palasudramanam S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT ceftazidime-resistant Klebsiella pneumoniae";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF467105; AAL75506.1; -
 DR InterPro: IPR001466; Beta_lactamase.
 DR Pfam: PF00144; beta-lactamase_1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR Plasmid.
 KW NON_TER

SEQUENCE 145 AA; 145 MW; DA95AF7557DA13D4 CRC64;
 SQ

SEQUENCE 145 AA; 15574 MW; F08634D6194B4C82 CRC64;
 SQ

Query Match 36.7%; Score 493; DB 2; Length 145;
 Best Local Similarity 68.1%; Pred. No. 7.4e-34;
 Matches 94; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

OY 100 AAMTSDNTANLTLTTGGPKELTAFLHNNGDHYTRLDRWPELENAIPNDEBRTMPV 159
 DB 7 AAXYMSDMSANLTLTTGGPKELTAFLHNNGDHYTRLDRWPELENAIPNDEBRTMPV 66
 OY 160 AAMTTRKLLTGLTGLTASRQQLIDMNEADRVAGPLRSALPAGWFIADKSGAGERSRG 219
 DB 67 SMAATRKLLTGLTGLTASRQQLIDMNEADRVAGPLRSALPAGWFIADKSGAGERSRG 126
 OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 263
 DB 127 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 144

RESULT 4

OY 052639 PRELIMINARY; PRT; 102 AA.
 AC 052639;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
 DE Beta-lactamase (Fragment).
 OS Pseudomonas aeruginosa.
 OC plasmid PRO1600.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TRANSPOSON-TnL;
 RA MEDLINE=94336757; PubMed=8058819;
 RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT "Deletion and transposon mutagenesis and sequence analysis of the
 RT prot600 orf1 region found in the broad-host-range plasmids of the pQF
 RT series.";
 RL Plasmid 31:265-274(1994).
 DR EMBL: L22691; AAA98312.1; -
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase_1.
 DR PRINTS: PR00118; BLACTAMASEA.
 DR Plasmid.
 KW NON_TER

SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;
 SQ

Query Match 36.3%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 1.2e-33;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AAMTTRKLLTGLTGLTASRQQLIDMNEADRVAGPLRSALPAGWFIADKSGAGERSRG 219
 DB 2 AAMTTRKLLTGLTGLTASRQQLIDMNEADRVAGPLRSALPAGWFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 263
 DB 62 IIAALGPDGKPSRIYVITTTGSOATMDERNROIAETGASLIKHW 99

RESULT 5

OY 0990F3 PRELIMINARY; PRT; 138 AA.
 AC 0990F3;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, last annotation update)

SEQUENCE 138 AA; 138 MW; DA95AF7557DA13D4 CRC64;
 SQ

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_30E_37S
Perfect score: 1345
Sequence: 1 HPEITLVKMDAEDQIGARVC.....TMDERNQIAEIGASLIKRW 263

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriophage:*
17: SP-archaeo:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result NO. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 524 | 39.0 | 105 | 2 | Q52026 pseudomonas |
| 2 | 524 | 39.0 | 105 | 2 | Q52330 escherichia |
| 3 | 493 | 36.7 | 145 | 2 | Q8RTD8 Klebsiella |
| 4 | 488 | 36.3 | 102 | 2 | Q52639 pseudomonas |
| 5 | 458 | 34.1 | 138 | 2 | Q99QF3 escherichia |
| 6 | 458 | 34.1 | 138 | 2 | Q9AM1 escherichia |
| 7 | 458 | 34.1 | 139 | 2 | Q9AM2 escherichia |
| 8 | 456 | 33.9 | 138 | 2 | Q9AM0 escherichia |
| 9 | 456 | 33.9 | 138 | 2 | Q9AM9 escherichia |
| 10 | 453 | 33.7 | 139 | 2 | Q9AM3 escherichia |
| 11 | 453 | 33.7 | 139 | 2 | Q9AM3 shigella fl |
| 12 | 338 | 25.1 | 67 | 2 | Q53553 Klebsiella |
| 13 | 320 | 23.8 | 95 | 2 | Q8GDES shigella fl |
| 14 | 316 | 23.5 | 62 | 2 | Q9JN58 shigella fl |
| 15 | 299 | 22.2 | 128 | 2 | Q8V000 escherichia |
| 16 | 245 | 18.2 | 180 | 2 | Q8KVT2 staphylococ |

| | | | | | |
|----|-------|------|-----|----|---------------------|
| 17 | 241 | 17.9 | 48 | 2 | Q9RLH0 proteus mir |
| 18 | 238 | 17.7 | 52 | 2 | Q9RA12 shigella fl |
| 19 | 197 | 12.6 | 38 | 2 | P97145 escherichia |
| 20 | 169.5 | 12.6 | 134 | 2 | Q8VUL3 staphylococ |
| 21 | 124.5 | 9.3 | 109 | 2 | Q53698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | P97146 escherichia |
| 23 | 87 | 6.5 | 100 | 2 | Q93505 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | Q9XB12 bacillus ce |
| 25 | 79 | 5.9 | 198 | 2 | Q9ACM8 streptococ |
| 26 | 77 | 5.7 | 181 | 16 | Q8D108 yersinia pe |
| 27 | 76.5 | 5.7 | 200 | 16 | Q97JEB clostridium |
| 28 | 76.5 | 5.7 | 202 | 16 | Q98520 rhizobium l |
| 29 | 75.5 | 5.6 | 184 | 5 | Q45379 caenorhabd |
| 30 | 74.5 | 5.5 | 195 | 16 | Q9RIAO streptococ |
| 31 | 74 | 5.5 | 153 | 5 | Q8S500 clone intes |
| 32 | 74 | 5.5 | 177 | 16 | Q9KEM7 bacillus ha |
| 33 | 73 | 5.4 | 175 | 17 | Q8ZTH8 pyrobaculum |
| 34 | 73 | 5.4 | 201 | 17 | Q8TW84 methanosarc |
| 35 | 72.5 | 5.4 | 113 | 2 | Q49970 mycobacteri |
| 36 | 72.5 | 5.4 | 145 | 2 | Q05984 staphylococ |
| 37 | 71.5 | 5.3 | 152 | 16 | Q9RI91 streptomyc |
| 38 | 71 | 5.3 | 131 | 2 | Q9X9H0 yersinia en |
| 39 | 71 | 5.3 | 169 | 17 | Q972W2 sulfobus |
| 40 | 71 | 5.3 | 192 | 16 | Q9HW07 pseudomonas |
| 41 | 71 | 5.3 | 203 | 4 | Q96BZ3 homo sapien |
| 42 | 70.5 | 5.2 | 131 | 16 | Q8YQL8 ambaba sp |
| 43 | 70.5 | 5.2 | 180 | 16 | Q9ZKK0 halobacte |
| 44 | 70 | 5.2 | 108 | 17 | Q9HR03 halobacteri |
| 45 | 70 | 5.2 | 150 | 10 | Q94IG4 nitocliana t |

ALIGNMENTS

RESULT 1

Q52026 PRELIMINARY: PRT: 105 AA.

AC Q52026; 01-NOV-1996 (TREMBL) 01, Created)
DT 01-NOV-1996 (TREMBL) 01, Last sequence update)
DT 01-MAR-2002 (TREMBL) 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN bla.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95011664; PubMed=7926843;
RA West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.;
RT *Construction of improved Escherichia-Pseudomonas shuttle vectors
RT derived from pUC18/19 and sequence of the region required for their
RT replication in Pseudomonas aeruginosa.";
RL Gene 148:81-86(1994).
RL EMBL: L30112; AAA66058.1; -.
DR HSSP: P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; Beta_lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON TER.
SQ SEQUENCE 105 AA: 11229 MW: D2889AA073330557 CRC64;

Query Match 39.0%; Score 524; DB 2; Length 105;

Best Local Similarity 100.0%; Pred. No. 1.1e-36;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATLRKLTGELTLASRQOLIDMWDKVGPIRLRSALPGWFTADKSGAGRGSRG 219
DB 2 AMATLRKLTGELTLASRQOLIDMWDKVGPIRLRSALPGWFTADKSGAGRGSRG 61

THIS PAGE BLANK (USPTO)

DB 134 SSDESPKEVOLA 145

RESULT 15

ID YDEI_ECOLI STANDARD; PRT: 172 AA.

AC P31131; 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein ydeI.

GN YDEI OR B1537.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=93186717; PubMed=8383113;

RA Cohen S.P., Haechler H., Levy S.B.;

RA *Genetic and functional analysis of the multiple antibiotic

RA resistance (mar) locus in Escherichia coli.*;

RA J. Bacteriol. 175:1484-1492(1993).

RT [2]

RN SEQUENCE FROM N.A.

RX MEDLINE=97426617; PubMed=9278503;

RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RA *The complete genome sequence of Escherichia coli K-12.*;

RA Science 277:1453-1474(1997).

RT [3]

RN SEQUENCE FROM N.A.

RX MEDLINE=97251357; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,

RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,

RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,

RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,

RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,

RA Yamamoto Y., Horikuchi T.;

RA *A 570-kb DNA sequence of the Escherichia coli K-12 genome

RA corresponding to the 28.0-40.1 min region on the linkage map.*;

RT DNA Res. 3:363-377(1996).

RL -I- SIMILARITY: BELONGS TO THE CINA FAMILY. STRONG, TO E.COLI YGAD.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

QY 73 SONDLVEYSPYREKHLTDGMYREICSAATMSDNTANLLTTIGP----- 120

DB 79 SOOSLERRYSAVSEK-----VAAEMATGAIERAD-ADVSTAITGCGFBGGEDGTPACTV 131

QY 121 -----KELTAFLHNMGDHYTRL 137

DB 132 WFAWHIKGQNTYAVMHFAGDCETVL 156

Search completed: September 10, 2003, 12:20:50
Job time : 5.57143 secs

Query Match 4.6%; Score 62.5; DB 1; Length 172;
Best Local Similarity 27.1%; Pred. No. 3.2e+02;
Matches 23; Conservative 9; Mismatches 26; Indels 27; Gaps 3;

```

Db      105  AALCECHQVGLVGDGLDELTKNLTADENATLELLKVLK 144
RESULT 13
HBG_RABIT STANDARD: PRT: 146 AA.
AC  P02099;
DT  21-JUL-1986 (Rel. 01, Created)
DT  21-JUL-1986 (Rel. 01, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Hemoglobin gamma chain (Beta-3).
GN  HBG.
OS  Oryctolagus cuniculus (Rabbit).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX  Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN  NCBI_TaxID=9986;
RP  SEQUENCE FROM N.A.
RX  MEDLINE=82053017; PubMed=6271761;
RA  Hardison R.C.;
RT  *The nucleotide sequence of rabbit embryonic globin gene beta 3.*;
RL  J. Biol. Chem. 256:11780-11786(1981).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=89178632; PubMed=2486295;
RA  Margot J.B., Demers G.W., Hardison R.C.;
RT  *Complete nucleotide sequence of the rabbit beta-like globin gene
RT  cluster. Analysis of intergenic sequences and comparison with the
RT  human beta-like globin gene cluster.*;
RL  J. Mol. Biol. 205:15-40(1989).
RN  [1]
RP  FUNCTION: THIS PROTEIN FUNCTIONS AS AN EMBRYONIC GLOBIN, BUT THE
CC  GENE STRUCTURE AND CHROMOSOMAL LOCATION RESEMBLE MORE CLOSELY THE
CC  HUMAN GAMMA CHAIN GENE, WHICH CODES FOR A FETAL GLOBIN.
CC  -1- SUBUNIT: Heterotrimer of two alpha chains and two gamma chains.
CC  -1- TISSUE SPECIFICITY: Red blood cells.
CC  -1- SIMILARITY: Belongs to the globin family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: M18818; AAA02984.1; -
DR  EMBL: V00883; CAA24252.1; -
DR  PIR: A02417; HBR3.
DR  HSSP: P02100; 1A9W.
DR  InterPro: IPR002337; Beta.haem.
DR  InterPro: IPR000971; Globin.
DR  Pfam: PF00042; globin.1.
DR  PRINTS: PRO0814; BETAHAEM.
DR  PROSITE: PS01033; GLOBIN.1.
KW  Heme; Oxygen transport; Transport; Erythrocyte; Embryo.
FT  INIT_MET 0
FT  SEQUENCE 146 AA; 16093 MW; EB8D6C1C24DD2D82 CRC64;
Query Match 4.6%; Score 62.5; DB 1; Length 146;
Best Local Similarity 30.8%; Pred. No. 2.6e+02;
Matches 32; Conservative 13; Mismatches 32; Indels 27; Gaps 6;

```

```

TATB_HELPY STANDARD: PRT: 160 AA.
ID  TATB_HELPY
AC  025700;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Sec-independent protein translocase protein tatB homolog.
GN  TATB OR HP1060.
OS  Helicobacter pylori (Campylobacter pylori).
OC  Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OX  Helicobacteraceae; Helicobacter.
RN  NCBI_TaxID=210;
RP  SEQUENCE FROM N.A.
RX  STRAIN=26695 / ATCC 700392;
RX  MEDLINE=97394467; PubMed=9252185;
RA  Tomb J.F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA  Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA  Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA  Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA  McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA  Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA  Cotton M.D., Meldrum J.M., Fujii C., Bowman C., Watley L., Wallin E.,
RA  Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA  Venter J.C.;
RT  *The complete genome sequence of the gastric pathogen Helicobacter
RT  pylori.*;
RL  Nature 388:539-547(1997).
RN  [1]
RP  FUNCTION: Required for correct localization of precursor proteins
CC  bearing signal peptides with the twin arginine conserved motif
CC  S/T-R-K-F-L-K. This sec-independent pathway is termed TAR for
CC  twin-arginine translocation system. This system mainly transports
CC  proteins with bound cofactors that require folding prior to export
CC  (by similarity).
CC  -1- SUBCELLULAR LOCATION: Inner-membrane bound (Probable).
CC  -1- SIMILARITY: Belongs to the tatB family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: AE000613; AAD08110.1; -
DR  PIR: D64652; D64652.
DR  TIGR: HP1060; -
DR  HAMAP: MF_00237; -; 1.
DR  InterPro: IPR003369; MtaA_Hcfl06.
DR  InterPro: IPR003998; TATB.
DR  Pfam: PF02416; MtaA_Hcfl06.1.
DR  PRINTS: PRO1506; TATBPROTEIN.
DR  TIGRFS: TIGR01410; tatB.1.
KW  Transport; Protein transport; Translocation; Transmembrane;
FT  Inner membrane; Complete proteome.
FT  TRANSMEM 1
FT  SEQUENCE 160 AA; 18332 MW; 3BD44234F7BABBD7 CRC64;
Query Match 4.6%; Score 62.5; DB 1; Length 160;
Best Local Similarity 20.5%; Pred. No. 2.9e+02;
Matches 27; Conservative 30; Mismatches 64; Indels 11; Gaps 5;

```

```

Db      35  FRSEERFPMASTFKVLLCGAVLSRIDAGDQIGRIHYSONDLVEYSPYTERKHLTDGMTV 94
3  FTAERKAATTSWK-----LVDEVDAGAAALGRL-----LVYI-FWQGRFDSFGNL 48

```

```

Db      19  FLGPEKEPQAVVDVKKFFRAVYKKTLDNARDTLDKEINIEIKETLEYQKLFENKVESLK 78
91  GMTVRELCSAATITMSDNTAANL-LTTTGKRELTAFLAHMCDHYTRLDRMEPELNAEI 148
79  GVKTEELDAKVT-AENEIKSTQIDLMQDYKSLFTWTIPNHLNEEVSN-----EALNKEV 133

```

RESULT 14

```

CC -----
DR EMBL: AE001728; AAD35606.1; -.
DR PIR: G72365; G72365.
DR HSSP: P11059; 1E94.
DR MEROPS: T01.006; -.
DR TIGR: TM0521; -.
DR HAMAP: MF_00248; -.
DR InterPro: IPR001353; Protease; 1.
DR Pfam: PF00227; Proteasome; 1.
KW Hydrolase; Protease; Complete proteome.
FT ACT_SITE 6
SQ SEQUENCE 176 AA; 18933 MW; ECCE369602A0ABD02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;
Best Local Similarity 28.2%; Pred. No. 3e+02;
Matches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

QY 117 ICGPKETAFILHNGDHVTRIDRMEPELNEALPNDERTTVPYAMATLRLKLTGELTL 176
DB 42 LGRGKVLACFAGSVADAMTLDFRFAKLR-----WGSNLTG 78
QY 177 ASGQQLDWMKADKAVAGLPSALPAGWFIADK-----SGAGERSGRTI-----ATL 224
DB 79 AAVELAKDW-RTDRV---LR-RLKALLVADKEMIFITISNGE-----VIGPDDDAAI 127
QY 225 GPDQ 228
DB 128 GSGG 131

RESULT 12
MARR_ECOLI STANDARD: PRT: 144 AA.
AC P27245; P76882; P77582;
DT 01-AUG-1992 (Rel. 23, created)
DT 16-OCT-2001 (Rel. 40, last sequence update)
DE 15-SEP-2003 (Rel. 42, last annotation update)
DE Multiple antibiotic resistance protein marr.
GN MARR OR SOXO OR CFXB OR INAR OR B1530.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID:562;
RN
RP SEQUENCE FROM N.A.
RA Cohen S.P., Haechler H., Levy S.B.;
RT "Genetic and functional analysis of the multiple antibiotic
RT resistance (marr) locus in Escherichia coli.";
RL J. Bacteriol. 175:1484-1492(1993).
RN
RP SEQUENCE FROM N.A.
RA STRAIN-K12 / MG1655;
RC MEDLINE-9742617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Zhao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE-97251357; PubMed-9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivaundaram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:365-377(1996).

```

```

RN [4]
RP CHARACTERIZATION.
RX MEDLINE-94110219; PubMed-8282690;
RA Ariza R.R., Cohen S.P., Bachhawat N., Levy S.B., Dimple B.;
RT "Repressor mutations in the marrAB operon that activate oxidative
RT stress genes and multiple antibiotic resistance in Escherichia
RT coli.";
RL J. Bacteriol. 176:143-148(1994).
RN
RP CHARACTERIZATION.
RX MEDLINE-95286534; PubMed-7768850;
RA Seoane A.S., Levy S.B.;
RT "Characterization of marr, the repressor of the multiple antibiotic
RT resistance (marr) operon in Escherichia coli.";
RL J. Bacteriol. 177:3414-3419(1995).
RN
RP MUTAGENESIS.
RX MEDLINE-20223625; PubMed-10760140;
RA Aleksun M.N., Kim Y.S., Levy S.B.;
RT "Mutational analysis of marr, the negative regulator of marrAB
RT expression in Escherichia coli, suggests the presence of two regions
RT required for DNA binding.";
RL Mol. Microbiol. 35:1394-1404(2000).
CC - FUNCTION: REPRESSOR OF THE MARRAB OPERON WHICH IS INVOLVED IN THE
CC ACTIVATION OF BOTH ANTIBIOTIC RESISTANCE AND OXIDATIVE STRESS
CC GENES. BINDS TO THE MARR OPERATOR/PROMOTER SITE.
CC - SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch)
CC -----
DR EMBL: M96235; AAC16394.1; ALT_INIT.
DR EMBL: AE000250; AAC74603.1; ALT_INIT.
DR EMBL: D90795; BAA15212.1; -.
DR EMBL: D90796; BAA15220.1; -.
DR EMBL: D90797; BAA15232.1; -.
DR PIR: E64807; E64907.
DR PDB: 1JGS; 28-DEC-01.
DR EcoGene; EG11435; marr.
DR InterPro: IPR000835; HTH_MARR.
DR Pfam: PF01047; MARR. 1.
DR PRINTS; PR00598; HTHMARR.
DR SMART; SM00347; HTH_MARR. 1.
DR PROSITE; PS01117; HTH_MARR_FAMILY. 1.
DR Transcription regulation; DNA-binding; Repressor;
KW Antibiotic resistance; Complete proteome; 3D-structure.
FT ->E: INCREASED TRANSCRIPTION OF THE
FT MUTAGEN 45 45
FT REGION II TRANSCRIPT.
FT NOTAGEN 77 77
FT R->L: INCREASED TRANSCRIPTION OF THE
FT MUTAGEN 123 144
FT MISSING: INCREASED TRANSCRIPTION OF THE
FT REGION II TRANSCRIPT.
SQ SEQUENCE 144 AA; 16065 MW; BE7DF5549E24D1D3 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 144;
Best Local Similarity 29.4%; Pred. No. 2.5e+02;
Matches 47; Conservative 14; Mismatches 46; Indels 53; Gaps 12;

QY 66 LGRRIHY--SQND--LVER-SP--VTEKILIDGMVRELCSAITMSDNTRANLLTLTG 118
DB 14 LGRILHVNOKRDLINELSPDLITAAOFKVLCSIR--CAACT----- 56
QY 119 GPKETAFILH-NMGDHVTRIDR-----WEPELNEALPNDERTTVPYAMATLRLKLTG- 171
DB 57 -PVELKVLVSDVLGALTRLDRLVCKGWVERLPN--PNDKRG-----VVKLTITGG 104
QY 172 -----ELTLASRQQLDWMKADKVA--GPLNSALP 201

```


CC EMBL: AE001079; AAB90876.1; -
 DR PIR: C69296; C69296.
 DR TIGR: AF0371; -
 DR HAMAP: MF_00241; -; 1.
 DR InterPro: IPR003265; Endo_3c.
 DR Pfam: PF00730; HNH-GPD; 1.
 DR SMART: SM00478; ENDO3C; 1.
 DR Hydrolase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
 KW Multifunctional enzyme; Complete proteome;
 FT ACT SITE 122
 SO SEQUENCE 198 AA; 22639 MW; 3A5C033A12F3FB CRC64;

Query Match 4.8%; Score 65; DB 1; Length 198;
 Best Local Similarity 24.6%; Pred. No. 2.4e+02;
 Matches 51; Conservative 25; Mismatches 85; Indels 46; Gaps 14;

QY 14 QLGARVGYIELDN---SGELLESFRSEPRPMSTFVLLCGAVLSRIDAGQ-----EQ 65
 DB 15 QLGKRG-GEVEFDFPFLDFSVKATIRTELAFICISANSATAGLKGRL-LGGVGYKEA 72
 QY 66 L---GRIHYSQNDLVESPTTEKHLTDGTVRELCSAATMSDNTAAILLTITG-GPK 121
 DB 73 LTLAGVRFH---NKKAEY-----IREAFKSFKLVEKALEASSKAREILLKIKGLGK 122
 QY 122 ELTAFALHNMG-DHYTRLD---RMEPELNEAIPNDERDTMPVNAATITRLKLL-----T 170
 DB 123 EASHFLRNAGREVAITIDRLKMLEROGYEPG---TMTAKRYLEVEKILMEISEER 177
 QY 171 GELLTLASROQLIDMMEADRVAGPLLR 197
 DB 178 GSEL---AEMDLRIWAE---MTGKVLK 198

RESULT 10
 RR4_PELNE STANDARD; PRT; 200 AA.
 ID RR4_PELNE
 AC 09FSD9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Chloroplast 30S ribosomal protein S4.
 GN RPS4.
 OS Pellia neesiana (Liverwort).
 OC Chloroplast.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
 CC Jungermanniopsida; Metzgeriidae; Fossombroniales; Pellineae;
 CC Pellaceae; Pellia.
 OX NCBI_TaxID=70144;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue-Cametophyte;
 RA Capesius J., Bloecher R.;
 RT "A molecular approach to bryophyte systematics."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: One of the primary rRNA binding proteins, it binds
 directly to 16S rRNA where it nucleates assembly of the body of
 the 30S subunit (By similarity).
 CC -1- FUNCTION: With S5 and S12 plays an important role in translational
 accuracy (By similarity).
 CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
 CC The interaction surface between S4 and S5 is involved in control
 of translational fidelity (By similarity).
 CC -1- SUBCELLULAR LOCATION: Chloroplast.
 CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@lsb-sib.ch.

CC EMBL: AJ250456; CAC14050.1; -
 DR HSSP: P81288; 1C05.
 DR HAMAP: MF_01306; -; 1.
 DR InterPro: IPR001912; Ribosomal_S4.
 DR InterPro: IPR002942; S4.
 DR InterPro: IPR005709; S4_bact_org.
 DR Pfam: PF00163; Ribosomal_S4; 1.
 DR Pfam: PF01479; S4; 1.
 DR SMART: SM00363; S4; 1.
 DR TIGRfam: TIGR01017; rpsD_bact; 1.
 DR PROSITE: PS00632; RIBOSOMAL_S4; 1.
 DR PROSITE: PS50889; S4; 1.
 KW Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
 FT DOMAIN 90
 SO SEQUENCE 200 AA; 23178 MW; 00D0098C6D9933F5 CRC64;

Query Match 4.7%; Score 63.5; DB 1; Length 200;
 Best Local Similarity 22.8%; Pred. No. 3.2e+02;
 Matches 39; Conservative 31; Mismatches 78; Indels 23; Gaps 7;

QY 14 QLGARVGYIE--LDLNSGELLESFRSEPRPMSTFVLLCGAVLSRIDAGQEQGRRIH 71
 DB 14 RLGALPGLSKTLESLGSGYSTPNKK---VSQYRIKL-----EKGKLRPH 58
 QY 72 Y---SONDLVEYSPYTEKHLTDGTVRELCSAATMSDNTAAILLTITGPKETLAFLE 128
 DB 59 YGLTERQLKTYRIANK--AKSGTGGLISQLEMRLDNIIFRLGMSPIPARQLVNHRR 116
 QY 129 NM-GDHYTRLDMEPELNEAIPNDERDTMPVNAATITRLKLLTGELLTLAS 178
 DB 117 ILINDNTVDIPSYNCEPKDVIYVNNKRESYIIKNDMSRKPKNHLLTFDS 167

RESULT 11
 HSLV_THEMA STANDARD; PRT; 176 AA.
 ID HSLV_THEMA
 AC 09WY21;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP-dependent protease hslv (EC 3.4.25.-).
 GN HSLV OR TM0521.
 OS Thermotoga maritima.
 CC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Colton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Elesen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 genome sequence of *Thermotoga maritima*."

CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
 CC complex (By similarity).
 CC -1- SUBUNIT: INTERACTS WITH HSLU (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@lsb-sib.ch.

DR HAMAP; MF_00003; -: 1.
 DR InterPro; IPR000238; Rib_bind_facta.
 DR Pfam; PF02033; Rbfa; 1.
 DR ProDom; PD007327; Rib_bind_facta; 1.
 DR TIGRFAMs; TIGR00082; rbfA; 1.
 DR PROSITE; PS01319; Rbfa; FALSE NEG.
 DR Trna_processing; Complete proteome.
 KM Sequence 116 AA; 13247 MW; A79DC71F0547514 CRC64;
 SQ
 Query Match 4.9%; Score 65.5; DB 1; Length 116;
 Best Local Similarity 28.9%; Pred. No. 1.1e+02;
 Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;
 OY 144 LNPAIPDERDTTPVAMATTLLRLKLTGEL-----LTLSAQQLIDWME-ADKVGAPLL 196
 DB 18 INNALANEINDKIKLAKLVAVR--LSNDLSYAKIFLDAHKRESMLKLVENKVSQ-L 74
 OY 197 RSLAPACW-----FIADKS 210
 DB 75 RSKLAEMTSYKVPDLRFVDET 97
 RESULT 8
 KIRY_METAC STANDARD; PRT; 206 AA.
 ID O8THS9;
 AC 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
 GN TMK OR M44433.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 NCBI_TaxID=2214;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C2A / ATCC 35395 / DSM 2834;
 RC MEDLINE=21929760; PubMed=11932238;
 RX Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., MacDonald P., Fitzhugh W., Calvo S., Engels R., Smitrov S., Althoor D., Brown A., Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R., Linton L., Mewson P., McKernan K., Talamas J., Tiroli A., Ye W., Zimmer A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M., Hoderich R., Ingram-Smith C., Kuelner H.C., Krzycki J.A., Guss A.M., Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K., Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C., Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I., Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E., Metcalf W.W., Birren B.;
 RA "The genome of Methanosarcina acetivorans reveals extensive metabolic and physiological diversity."
 RT Genome Res. 12:532-542(2002).
 RL
 CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine 5'-diphosphate.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC
 CC EMBL; AE011164; AM07774.1; -.
 DR HAMAP; MF_00165; -: 1.
 DR InterPro; IPR000062; Thymidylate_kin.
 DR Pfam; PF02223; Thymidylate_kin; 1.
 DR TIGRFAMs; TIGR00041; dTMP_kinase; 1.
 DR PROSITE; PS01331; THYMIDYLATE_KINASE; FALSE NEG.
 DR Trna_processing; Complete proteome.
 KM Complete proteome.
 KW Complete proteome.

FT NP_BIND 10 17 ATP (POTENTIAL).
 SQ SEQUENCE 206 AA; 23303 MW; A19C2E5B8C5423B3 CRC64;
 Query Match 4.9%; Score 65.5; DB 1; Length 206;
 Best Local Similarity 18.7%; Pred. No. 2.3e+02;
 Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;
 OY 56 LSRID-AGOEQJGRRIHYSONDLVEYSPVTEKHITDGTWVRELCSAATMSQNTANLL 114
 DB 8 LRGIDSSGKSTYAKKID-ENSELRFEPVFTRETRG-----TLTGNAYENALQ 55
 OY 115 TTIGGPKELTAFLNNMDHYTRLDWMEPELNEAIPNDERDTTPVAMATTLLRLKLTGELL 174
 DB 56 SDTDLAELEFLFADNHEHLAKLYKPALEDGKTYISDRYSOSRYAVOGITLKRDLNPL- 114
 OY 175 TLASROQLIDWMEADKVGAPLLNSALPAGW-----FIAD-----KSG-AGE----- 214
 DB 115 -----EMVR-----DLHGMVFIPIDLFLFDIEPIAVKRCGRGQTRF 154
 OY 215 -----RGRGIIAALGPDKPSRIIVYITGSOATME 247
 DB 155 EKIEPLRGVRELFGLAE-EPERFVIADAGSPEDVER 192
 RESULT 9
 OGGL_ARCFU STANDARD; PRT; 198 AA.
 ID OGGL_ARCFU
 AC 029876;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable N-glycosylase/DNA lyase (includes: 8-oxoguanine DNA glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18) (AP lyase)).
 GN OGG OR AF0371.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
 OC Archaeoglobaceae; Archaeoglobus.
 NCBI_TaxID=2234;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RC MEDLINE=98049343; PubMed=9389475;
 RX Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D., Richardson D.L., Kachlavage A.R., Graham D.E., Kyriades N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.T., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.D., Spriggs T., Arllich P., Kalne B.P., Sykes S.M., Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus."
 RT Nature 390:364-370(1997).
 RL
 CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of guanine (7,8-dihydro-8-oxoguanine = 7-oxoG) from DNA. Also nicks DNA at apurinic/apyrimidinic sites (AP sites) (by similarity).
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or apyrimidinic site in DNA is broken by a beta-elimination reaction, leaving a 3'-terminal unsaturated sugar and a product with a terminal 5'-phosphate.
 CC -1- SIMILARITY: BELONGS TO THE OGGL FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).
 CC


```

OY 44 MSFFKVLGCAVLSRIDAGQEQRLRRIRYSNDLV-EYSP-VTEKHILTDGMTVRELCSAA 101
DB 1 MKS1KVCVVG-----EGGIGKTSMLSTYSNSISNEQPTVFDNST----- 42
OY 102 ITMSDNTAANLLTTTIGGPKELTAF-----LHNMCDHVTRLDMWEDELN 145
DB 43 LHMHNKRPYNLSLMDTAGEEFSKILRLSYPTDVFLLFSLINPSSFNILDSWQELN 102
OY 146 EAIPNDERDFTMPVAMATLRLKLLNGELL--TLASHQOL 182
DB 103 ENCPN-----TPYLVGTQMDLKSNSVILDRICEKQOL 135

RESULT 4
HC3L_THIFE STANDARD; PRT; 122 AA.
AC P80509;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
DE (Fragments).
GN HOXG.
OS Thiodacillus ferrooxidans.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
OC Acidithiobacillaceae; Acidithiobacillus.
OX NCBI_TaxID=920;
RN [1]
RP SEQUENCE.
RC STRAIN=ATCC 19859; PubMed-8661919;
RX MEDLINE=96241862; PubMed-8661919;
RA Fischer J., Quenemeler A., Kostka S., Kraft R., Friedrich C.G.;
RT "Purification and characterization of the hydrogenase from
RT Thiodacillus ferrooxidans."
RL Arch. Microbiol. 165:289-296(1996).
CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 = 2 H(+) + 2
CC ferrocyclochrome c3.
CC -1- COFACTOR: IRON.
CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
CC ALSO UNKNOWN IF CAPS EXIST BETWEEN THE PEPTIDES.
KW Oxidoreductase; Iron.
FT NON_CONS 20 21
FT NON_CONS 29 30
FT NON_CONS 35 36
FT NON_CONS 42 43
FT NON_CONS 59 60
FT NON_CONS 72 73
FT NON_CONS 78 79
FT NON_CONS 87 88
FT NON_CONS 98 99
FT NON_CONS 107 108
FT NON_TER 122 122
SO SEQUENCE 122 AA; 13146 MW; EB4PAF365348DA0B CRC64;

Query Match 5.1%; Score 68; DB 1; Length 122;
Best Local Similarity 26.4%; Pred. No. 74;
Matches 28; Conservative 14; Mismatches 22; Indels 42; Gaps 4;

OY 19 VOYIELDLMSGELLESFRSERRPMSTFKVLLCGAVLSRIDAGQEQRLRRIRYSNDLV 78
DB 12 VGRVEGDLDSXSLIEFRN-----AIIARFGCG-----LG 41
OY 79 EYSPVTEKHILTDGMTVRELCSAAITMSDNTAANLLTTTIGGPKELT 124
DB 42 KYAPFGTNYETIGVT-----ISGDKPDQAGLVVT-----PREST 75

RESULT 5
ALKH_BACSU STANDARD; PRT; 196 AA.
AC P50846;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)

```

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE KHG/KDPG aldolase [Includes: 4-hydroxy-2-oxoglutarate aldolase
DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KHG-aldolase); 2-
DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (Phospho-2-
DE dehydro-3-deoxygluconate aldolase) (phospho-2-keto-3-deoxygluconate
DE aldolase) (2-keto-3-deoxy-6-phosphogluconate aldolase) (KDPG-
DE aldolase)].
GN KPGA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;
RX MEDLINE=96349105; PubMed-8760912;
RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the sera and kdg loci cloned in a yeast artificial chromosome."
RL Microbiology 142:2005-2016(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulle M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Meuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelie D., Porwollik S., Prescott A.M.,
RA Presseau E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivoita C., Rochna E., Roche B., Rose K., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartotil A.,
RA Viari A., Wandut R., Wedler E., Wedler H., Weitenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate = pyruvate +
CC glyoxylate.
CC -1- PATHWAY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate +
CC pyruvate + D-glyceraldehyde 3-phosphate.
CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
CC OF GLYOXYLATE.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE KHG/KDPG ALDOLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 5.57143 Seconds
(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E_37S

Profile score: 1345
Sequence: 1 HPEPTLVKVKDAEDQLGARVC.....TMDERNRQIAETGASLIRKM 263

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|-----------------------|
| 1 | 78 | 5.8 | 191 | Y0A4_SCHPO | Q10346 echinosacch |
| 2 | 69.5 | 5.2 | 196 | RAC2_LOTJA | Q40220 lotus japon |
| 3 | 69 | 5.1 | 201 | RACG_LOTJA | Q99800 dictyosteli |
| 4 | 68 | 5.1 | 122 | HC31_THIFE | P80509 thibacilli |
| 5 | 66.5 | 4.9 | 196 | ALKH_BACSU | P50846 b.kh/kdpy |
| 6 | 66 | 4.9 | 149 | NDK_TREPA | O83974 treponema p |
| 7 | 65.5 | 4.9 | 116 | RBP4_VREPA | O83940 ureaplasma |
| 8 | 65.5 | 4.9 | 206 | KTHY_METAC | O81859 methanosarc |
| 9 | 65 | 4.8 | 198 | OGG1_ARCFU | O29876 archaeoglob |
| 10 | 63.5 | 4.7 | 200 | RR4_PELNE | O9f649 pellicula nees |
| 11 | 63 | 4.7 | 176 | HSIV_THEMA | O83974 treponema p |
| 12 | 62.5 | 4.6 | 144 | MARR_ECOCI | P27245 escherichia |
| 13 | 62.5 | 4.6 | 146 | HBC_RABIT | P02099 oryctolagus |
| 14 | 62.5 | 4.6 | 160 | TANB_HELIPY | O25700 helicobacte |
| 15 | 62.5 | 4.6 | 172 | YDEI_ECOCI | P11131 escherichia |
| 16 | 62.5 | 4.6 | 182 | PYRE_STRCO | O9x847 streptomyce |
| 17 | 62.5 | 4.6 | 184 | HRLP_PSESY | P37929 pseudomonas |
| 18 | 62.5 | 4.6 | 195 | TRPF_THEVO | O87946 thermoplasma |
| 19 | 62 | 4.6 | 173 | Y265_BORBU | O51280 borrelia bu |
| 20 | 62 | 4.6 | 174 | IHBH_RAT | P17491 rattus norv |
| 21 | 62 | 4.6 | 182 | RR4_BELCH | O83950 belamcanda |
| 22 | 62 | 4.6 | 182 | Y861_BIFLO | O83950 belamcanda |
| 23 | 61.5 | 4.6 | 152 | Y033_YERPE | O83950 belamcanda |
| 24 | 61.5 | 4.6 | 184 | MUR1_SCHPO | O9uug5 schistosacch |
| 25 | 61.5 | 4.6 | 185 | YCJC_ECOCI | P38522 escherichia |
| 26 | 61.5 | 4.6 | 197 | HANI_XENLA | O73615 xenopus lae |
| 27 | 61 | 4.5 | 177 | VNSC_XENLA | P35948 xenopus lae |
| 28 | 61 | 4.5 | 177 | VNSC_RINDR | O03333 rinderpest |
| 29 | 61 | 4.5 | 184 | ARL2_DROME | O06849 drosophila |
| 30 | 61 | 4.5 | 194 | RR4_TRIEN | O20234 iris ensata |
| 31 | 60.5 | 4.5 | 126 | YGM1_YEAST | P53170 saccharomyc |
| 32 | 60.5 | 4.5 | 142 | YEG3_METTM | O50770 methanobact |
| 33 | 60.5 | 4.5 | 150 | PDOV_SALTY | O9x406 salmonella |

| | | | | | | |
|----|------|-----|-----|---|-------------|--------------------|
| 34 | 60.5 | 4.5 | 160 | 1 | TATB_HELIPY | Q92m58 helicobacte |
| 35 | 60.5 | 4.5 | 178 | 1 | HSIV_RALSO | O8y3d7 ralatonia s |
| 36 | 60 | 4.5 | 121 | 1 | SECR_HUMAN | P09683 homo sapien |
| 37 | 60 | 4.5 | 178 | 1 | PYRE_ARCFU | O28533 archaeoglob |
| 38 | 60 | 4.5 | 178 | 1 | UCRI_ANASP | P70758 anabaena sp |
| 39 | 60 | 4.5 | 200 | 1 | TATB_CAUCR | O9a6t1 caulobacter |
| 40 | 59.5 | 4.4 | 103 | 1 | R510_NEIGO | P48851 neisseria g |
| 41 | 59.5 | 4.4 | 156 | 1 | BPR_AZOVY | P22759 azotobacter |
| 42 | 59.5 | 4.4 | 177 | 1 | ATPD_HAEIN | P43717 haemophilus |
| 43 | 59.5 | 4.4 | 184 | 1 | RL6_MYCPN | O50303 mycoplasma |
| 44 | 59.5 | 4.4 | 193 | 1 | SAR2_LYCES | P52884 lycopersico |
| 45 | 59.5 | 4.4 | 198 | 1 | UPP_ARATH | O65583 arabidopsis |

ALIGNMENTS

RESULT 1
YD44_SCHPO STANDARD; PRT: 191 AA.
AC Q10346;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1F12.04c in chromosome I.
GN SPAC1F12.04c.
OS Schistosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schistosaccharomycetes;
OC Schistosaccharomycetales; Schistosaccharomycetaceae;
OC Schistosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gallilium R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks R., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Gobie A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicakeat G., Aert R., Robben J., Grymoprez B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gebel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Horzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Roche M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Love T., McCombie W.R., Paulsen I., Potashkin J.,
RA Sipkavski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schistosaccharomyces pombe."
RL Nature 415:871-880(2002).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).

DR EMBL: Z69944; CAAS3808.1; -
DR PIR: T38062; S67447.
DR GeneDB_Spombe; SPAC1F12.04c; -

THIS PAGE BLANK (USPTO)

```

Oy      148 IPNDEROTMPVAMATTLRKLLTGLLT 175
          : :  | | | : : | : |
Db      113 LPGETYSVTVPFGSAPTRVKKVTTGTGVT 140

```

RESULT 14

DB87360
 conserved hypothetical protein CC0895 [imported] - *Caulobacter crescentus*
 C:Species: *Caulobacter crescentus*
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: DB87360
 R:NIEMANN, W.C.; FELDBLYUM, T.V.; PAUSEN, I.T.; NELSON, K.E.; EISEN, J.; HEIDELBERG, J.
 B.; LAUB, M.T.; DOBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWIM, M.L.; HAFT, D.H.; KOLMAN,
 N. J.; ERMOLAEVA, M.; WHITE, O.; SALZBERG, S. L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.
 A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A:Reference number: AB7249; MIMD:21173698; PMID:11259647
 A:Accession: DB87360
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-167 <S>
 A:Cross-references: GB:AE005673; NID:G13422160; PIDN:AAK22880.1; GSPDB:GN00148
 C:Genetic: CC0895

Query Match 5.2%; Score 69.5; DB 2; Length 167;
Best Local Similarity 25.4%; Pred. No. 1.4e+02;
Matches 36; Conservative 19; Mismatches 56; Indels 31; Gaps 6

[illegible]

RESULT 15

AD350b
malybdopterin biosynthesis mog protein [Imported] - *Brucella melitensis* (strain 16M)
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AD3606
R:DelVecchio, V.G.; Kapralov, V.; Redkar, R.U.; Patra, G.; Mijer, C.; Los, T.; Ivanova, R.; Delvecchio, V.G.; Kapralov, V.; Redkar, R.U.; Patra, G.; Mijer, C.; Los, T.; Ivanova, R.; Mazur, M.; Gotsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Detess, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A:Reference number: AD3352; PMID:1175658

| | | | | |
|-----------------------|--------------|-------------|----------------|-------------|
| Query Match | 5.1% | Score 68.5: | DB 2: | Length 172: |
| Best Local Similarity | 24.8% | Pred. No. 1 | 7e+02: | |
| Matches 39: | Conservative | 14: | Mismatches 43: | Indels 61: |
| | | | | Gaps 9: |

```

OY      83 VTKEHLTDGN--TVRELCSSAATITNSDNTAANLLTTTIG--GPRELTFEINMGCHVTRIDM 140
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      39 TVRRVIPDGMESYVD---TLIDCDTYACDLILTTGGTGS----- 76

OY      141 EPELNAIPDPDERDITMPVAMATTLRKLLT--GELLTJLASRQQLIDMHEADKVAGELLS 198

```


[illegible]

RESULT 11
 C71869
 hypothetical protein jhp0935 - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: C71869
 R:Alm, R.A.; Ling, U.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jlang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho-
 A:Reference number: A71800; M0ID:99120357; PMID:9923682
 A:Accession: C71869
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-180 <ARN>
 A:Cross-references: GB:AE001523; GB:AE001439; NID:g4155513; PIDN:AAD06513.1; PID:g415551
 A:Experimental source: strain J99
 A:Genetics:
 A:Gene: jhp0935

| | | | | |
|--------------------------|--------|--------------------|------------|-------------|
| Query Match | 5.28; | Score 70.5; | DB 2; | Length 180; |
| Best Local Similarity | 25.48; | Pred. No. 1.2e+02; | | |
| Matches 47; Conservative | 28; | Mismatches 43; | Tndels 67; | Cover 10 |

```

OY 30 EILSESRSEBRFPMMSTFKVLLCGAVLSRIDAGQEOIGRIHYSONDV---EYSPYTE 85
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 2 EFPATIRAEKEREPFSLF-----NRSSGSPDLKQWKSKEYENILLDTKGSYKETO 52
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 86 KHLTDMYVELCSAATMSDNFANLLL--TTIGPKEL-YAFLHMGCHVTRLDREMP 142
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 53 K-----AMLS-----NYLVPTT--PSOLDDEVLANMLERIEQIQ----- 86
    | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
OY 143 ELNE---AIPNDRDPTMPVAMATTLKRLTGLGELLTASROOIIDME---ADKVAGPL 195
    | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 87 ELNENRALIVIMRPTIP-----TLERQDLIEETIKENPSPDRIT--L 128
    | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

```

| | | | |
|----|-----|-------|-----|
| QY | 196 | LRSAI | 200 |
| | | 11:1 | |
| Db | 129 | LESSL | 133 |

RESULT 12
 EB4217
 hypothetical protein Vng0594h [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: EB4217
 R:Ng, M.Y.; Kennedy, S.P.; Mahairs, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky,
 J.; Lelthausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Je-
 Jung, K.H.; Alam, M.; Freltas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omar, A.D.; Edhardt, H.; Lowe, T.M.;
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; NCID:20504483; PMID:11016950
 A:Accession: EB4217
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-108 <STO>
 A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AAIG19105.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG0594H

| | | | | |
|-------------|--------|---------------|-------|----------------|
| Query Match | 5.28; | Score 70; | DB 2; | Length 108; |
| Best Local | 27.58; | Pred. No. 68; | | |
| Matches | 28; | Conservative | 16; | Mismatches 42. |
| | | | | Totals 15. |

QY 44 MSPTKVLGCVNLRIDAGOEOLRRIRIYSONDVEYSPTVETKHLDGTATVELCSAII 103
| | | | | : : : : :
| | | | | : : : : :
Db 1 MDITVTELDGALLRLAEATD-----RVEEVRPDALFVTDTLRFRRHDSRV-----GSII 50

```

      | | : | : | | : | | |
Db   51 NDDGTDRTMARTVPGDSDFIAVEPTSFVAIVDATRTDR 92

```

RESULT 13
 F84251
 hypothetical protein Vng0953c [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84251
 R:Ng, W.V.; Kennedy, S.P.; Mahairs, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky,
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Fretz, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.;
 A:title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: F84251
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-145 <STO>
 A:Cross-references: GB:AE004437; NID:g10580510; PIDN:AAG19378.1; GSPDB:GND0138
 C:Genetics:
 ;:Gene: VNG0953C

| | | | | |
|-----------------------|--------|--------------------|--------|----------------|
| Query Match | 5.28; | Score 69.5; | DB 2; | Length 145; |
| Best local Similarity | 20.38; | Pred. No. 1.1e+02; | | |
| Matches | 30; | Conservative | 29; | Mismatches 66; |
| | | | Indels | 23; |
| | | | Gaps | 3 |

QY 41 EPANSTFVLVLCGAVLS-----RIDACGEOLGRRIHSQNDLVETSPYTEKH 87
 Db 3 FSVSGSAALFLPAAVSVGILSMAFNGEGERQDANNARNDKRLAKTNTVEAANTYDS 62
 QY 88 LTDDMTATRELCSAATMSDNTAANLLLTITGCPKELTAFLAHMGCHVTRLDRWEPELNEA 147
 Db 63 VNDIVTATNATNGSTLSVSQTDVLDV-----DGYVTDSAAVSSSVYDGNSTQDIDM----- 112

[illegible][illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E_37S
Perfect score: 1345
Sequence: 1 HPETLVKVDKADQLGARVG.....TMDERNROIAGSLIKHM 263

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

1: PIR_76:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 39.0 | 105 | 2 | JC2566 |
| 2 | 78 | 5.8 | 191 | 2 | S67447 |
| 3 | 76.5 | 5.7 | 200 | 2 | G97064 |
| 4 | 75.5 | 5.6 | 184 | 2 | T21126 |
| 5 | 74.5 | 5.5 | 195 | 2 | T36975 |
| 6 | 74 | 5.5 | 177 | 2 | A83753 |
| 7 | 72.5 | 5.4 | 113 | 2 | T45195 |
| 8 | 71.5 | 5.3 | 152 | 2 | T36984 |
| 9 | 71 | 5.3 | 192 | 2 | G83096 |
| 10 | 70.5 | 5.2 | 131 | 2 | AD2281 |
| 11 | 70.5 | 5.2 | 180 | 2 | C71869 |
| 12 | 70 | 5.2 | 108 | 2 | F84217 |
| 13 | 69.5 | 5.2 | 145 | 2 | F84251 |
| 14 | 69.5 | 5.2 | 167 | 2 | D87360 |
| 15 | 68.5 | 5.1 | 172 | 2 | AD3606 |
| 16 | 68.5 | 5.1 | 190 | 2 | T09136 |
| 17 | 68 | 5.1 | 170 | 2 | AB0192 |
| 18 | 68 | 5.1 | 177 | 2 | T48420 |
| 19 | 67.5 | 5.0 | 192 | 2 | A83587 |
| 20 | 67.5 | 5.0 | 198 | 2 | D95285 |
| 21 | 66.5 | 4.9 | 128 | 2 | F87353 |
| 22 | 66.5 | 4.9 | 188 | 1 | G69365 |
| 23 | 66.5 | 4.9 | 196 | 2 | H69647 |
| 24 | 66 | 4.9 | 146 | 2 | C72703 |
| 25 | 66 | 4.9 | 149 | 2 | F71252 |
| 26 | 65.5 | 4.9 | 42 | 2 | F56978 |
| 27 | 65.5 | 4.9 | 116 | 2 | C82906 |
| 28 | 65.5 | 4.9 | 195 | 2 | AE0623 |
| 29 | 65.5 | 4.9 | 197 | 2 | H90211 |

| | | | | | | |
|----|------|-----|-----|---|--------|--------------------|
| 30 | 65 | 4.8 | 148 | 2 | E75283 | conserved hypothet |
| 31 | 65 | 4.8 | 150 | 2 | AC0284 | probable heat choc |
| 32 | 65 | 4.8 | 153 | 2 | E81708 | conserved hypothet |
| 33 | 65 | 4.8 | 160 | 1 | E69186 | conserved hypothet |
| 34 | 65 | 4.8 | 168 | 2 | B75498 | conserved hypothet |
| 35 | 65 | 4.8 | 180 | 2 | G70912 | hypothetical prote |
| 36 | 65 | 4.8 | 198 | 2 | C69296 | conserved hypothet |
| 37 | 64.5 | 4.8 | 151 | 2 | D81333 | probable protein-t |
| 38 | 64.5 | 4.8 | 177 | 2 | D90227 | hypothetical prote |
| 39 | 64 | 4.8 | 113 | 2 | D70580 | hypothetical prote |
| 40 | 64 | 4.8 | 150 | 2 | T08585 | calmodulin - soybe |
| 41 | 64 | 4.8 | 162 | 2 | AG0769 | probable acetylira |
| 42 | 64 | 4.8 | 168 | 2 | C85715 | unknown protein en |
| 43 | 64 | 4.8 | 197 | 2 | AC2133 | iron(iii) diclrat |
| 44 | 63.5 | 4.7 | 168 | 2 | T20806 | hypothetical prote |
| 45 | 63.5 | 4.7 | 179 | 2 | AB1994 | hypothetical prote |

ALIGNMENTS

RESULT 1

JC2566
bla protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa
C:Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996
C:Accession: JC2566
R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 14811281, 81-86, 1994

A:Title: Construction of improved Escherichia-Pseudomonas shuttle vectors derived fro
A:Reference number: JC2565
A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566
A:Molecule type: DNA
A:Residues: 1-105 <WES>

C:Genetics:
A:Gene: bla

C:Superfamily: beta-lactamase I

Query Match 39.0%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 9.9e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| OY | 160 | AMATTLRLKLTGELLTSLASROQLIMMEADKYAGPLRLSALPGWFTADSGAGERSRG | 219 |
| Db | 2 | AMATTLRLKLTGELLTSLASROQLIMMEADKYAGPLRLSALPGWFTADSGAGERSRG | 61 |
| OY | 220 | IIAALGPDGKPSRIYVITTSQATMDERNROIAGSLIKHM | 263 |
| Db | 62 | IIAALGPDGKPSRIYVITTSQATMDERNROIAGSLIKHM | 105 |

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: T38062; S67447
R:McLean, J.; Harris, D.; Bartrell, B.G.; Rejandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996

A:Reference number: T38062
A:Accession: T38062
A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:269944; NID:q1217974; PIND:CAA93808.1; PIND:q1217978; GSPDB:G

A:Experimental source: strain 972h-; cosmid c1F12

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.8%; Score 78; DB 2; Length 191;

```

RESULT 15
US-10-106-698-4858
; Sequence 4858, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106, 698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4858
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4858

```

```

Query Match          4.88; Score 65; DB 15; Length 179;
Best Local Similarity 27.38; Pred. No. 1.2e+02;
Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;

QY 120 PKELAFALHNGVHTRLD---RMEPELNEA-----IPDDEOTTPVAMATTLRLKLTG 171
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 52 PRLIIV---QGEKLEVLDSKRWMLVYNEAGRSGLIPSNILPLDPTGTGQCSRSRV 108
   | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 172 ELITLASR-QQLIDMMEADKRVAGPLLR 198
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 109 PMRLRSSRPEEVDWDLQAEENFSTATVPT 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: September 10, 2003, 12:33:19
Job time : 18.5714 secs

Best Local Similarity 20.9%; Pred. No. 85;
Matches 39; Conservative 26; Mismatches 68; Indels 54; Gaps 7;
QY 50 LCCAVLSRIDAGQOLGRIRHYSOND--LVEXSPYTEKHL-----TDGMYRDL----- 97
Db 17 LLLAAVLRTOADVPVPRATRLPVEAKDCHIAQFKSISPELQAFKKAQIEKLEKDM 76
QY 98 -CSA-----AIFMS--DNTANLLTTIGGPKELTAFLN 129
Db 77 RCSSHLISRANDLQLOQVOERPKALOAEVALTLKWMENINDSALTITLQPLHTLSHS 136
QY 130 MGDHVT-----RLDRWELNEAIPNDRDFTMPVAMATTIRKLTGELLTL 176
Db 137 QLOCTQLOATAEPKPPSRRLSRWLRLQDA-QSKETPGCLEDSVSNLQLLRLDLKCV 195
QY 177 ASROOLI 183
Db 196 ASDQCV 202

RESULT 12
US-09-764-870-390
; Sequence 390, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764, 870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-870-390

Query Match 4.9%; Score 65.5; DB 9; Length 191;
Best Local Similarity 19.7%; Pred. No. 1.1e+02;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;
QY 21 YIELDNGSELIESRSEERFPMSSTFKVLLCGAVLSRIDAGQOLGRIRHYSONDLVEY 80
Db 48 YCRSLINDSEV-----DMFGDYDSFTENSFIQVDDLEOK-----YMQ----- 85
QY 81 SPVTEKHLTDGMTVRELCSAAITMSDNTANLLTTIGGPRELTAFLH----- 128
Db 86 LPEKKHATDTRAT-ENLCSESI-----KNKLSITTTIGNLTQLTKHTENOSGEGVTI 138
QY 129 -----NMGDHVTRLDWEPR-----LNEAIPND 151
Db 139 EPGADLLYVPSSQAIFYXNLQNSSNDLDGHSKMDRWXSSSHNTVNEELPHN 191

RESULT 13
US-10-125-540-390
; Sequence 390, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PT214C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc-feature
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc-feature
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-125-540-390

Query Match 4.9%; Score 65.5; DB 15; Length 191;
Best Local Similarity 19.7%; Pred. No. 1.1e+02;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;
QY 21 YIELDNGSELIESRSEERFPMSSTFKVLLCGAVLSRIDAGQOLGRIRHYSONDLVEY 80
Db 48 YCRSLINDSEV-----DMFGDYDSFTENSFIQVDDLEOK-----YMQ----- 85
QY 81 SPVTEKHLTDGMTVRELCSAAITMSDNTANLLTTIGGPRELTAFLH----- 128
Db 86 LPEKKHATDTRAT-ENLCSESI-----KNKLSITTTIGNLTQLTKHTENOSGEGVTI 138
QY 129 -----NMGDHVTRLDWEPR-----LNEAIPND 151
Db 139 EPGADLLYVPSSQAIFYXNLQNSSNDLDGHSKMDRWXSSSHNTVNEELPHN 191

RESULT 14
US-09-764-868-757
; Sequence 757, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764, 868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 757
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-757

Query Match 4.8%; Score 65; DB 10; Length 179;
Best Local Similarity 27.3%; Pred. No. 1.2e+02;
Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;
QY 120 PRELTAFLHNMGDHVTRLD-----RWPELANA-----IPNDRDFTMPVAMATTIRKLLTG 171
Db 52 PRELTVV--GGEKLEVLDSKRWMLVKNAGRGYIPSNILEPQPGTGTGOSPSRY 108
QY 172 ELUTLASR-QQLDWMEDKVNAGPLRS 198
Db 109 PMLRLSRPEEYTDWLOAENFSTATVAT 136

```

; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-765
```

```

Query Match          5.0%; Score 67.5; DB 15; Length 184;
Best Local Similarity 24.7%; Pred. No. 66;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;
```

```

QY 41 PPMSTFKVL-----LCGAVLSRIDAGOEQLGRRIHYSQNDLVEYSPV---TEKHLTDG 91
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 23 FCMSTERLLVYPYMLNGSVASRL-----RDSINGKPALDMPTRKRISIG 67
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 92 -----MTVREICSAITMSDNTAANLL-----TTIGPKELTAFLHMGDHYTRLDNR-- 140
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 68 AARGLYLHEQCCKPKIIRVKANILLDEYFEAVVGGFGLAKLLDHRDHSHTTAVRGTV 127
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 141 --EPELNEAIPNDERDTPVAMATLRLKLTGEL---LTLASROQ---LIDPM 186
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 128 GHIAEYILSTGQSSER--TDFPGILLLEITGOKALDFGAANQKVALDMV 179
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```

RESULT 9
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: Theell et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-948-018-8
```

```

Query Match          5.0%; Score 67.5; DB 10; Length 187;
Best Local Similarity 22.4%; Pred. No. 68;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;
```

```

QY 112 LLLTTIGPKELTAFLHMGDHYTRLDNRWEPE-----LNEAIPNDERDTPVAMAT- 163
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 11 LLLLAAPPERS-----QYCGRLTYWPDNMCSSCLDPRFGP-----PPCGALETG 58
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 164 -----TURKLTGELLTLASR---OQIDWMEA-----DKVAGP-----L 195
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 59 DTKRESLPLSLRSLASPLSLRLDELVELELLVLDPRPGGGMAGTTRHILA 118
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 196 LRSALPACGFADKSGAGERSGCIITALGPDGRPSR--IVVITYTGSQATDNRNRQIA 253
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 119 ARYGLPAAW-----STFAVSLRPSRSPPLRLIEVYVAREPSASISGQLCTHILA 165
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 254 EIG 256
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 166 QLG 168
```

```

RESULT 10
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMEGA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIRAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623
```

```

Query Match          5.0%; Score 67; DB 15; Length 194;
Best Local Similarity 38.6%; Pred. No. 81;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;
```

```

QY 202 AGWFIADKSGAGERSGNG-----ITAAIGPDG--KPSRIYVITTT 239
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 75 SGRRVTPRSGAGERYTEGAERTYTAIGPLAVHEPRVAVVAVVT 118
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

```

RESULT 11
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Shepard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klueber, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kinsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-11
```

```

Query Match          5.0%; Score 67; DB 15; Length 202;
```

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8136
LENGTH: 149
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8136

Query Match 5.1%; Score 69; DB 15; Length 149;
Best Local Similarity 30.9%; Pred. No. 34;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;

QY 168 LTTGGKELTAFLHNGDHH--VTRIDRMPELNEAIPNDRDTPVAMATTLRLKLT 211
DB 50 LLDIELTFLKRLVAVSDYDAKAKEMGIDWHEHD---PALSSHADGGELAEENRLREI 105
QY 212 AGEGSGRTIATLPDGRPSR 232
DB 106 AGLDAAQLPSEGPAPER 126

RESULT 6
US-09-738-626-5680
Sequence 5680, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIYOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIOI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MACHO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5680
LENGTH: 178
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680

Query Match 5.1%; Score 69; DB 10; Length 178;
Best Local Similarity 22.1%; Pred. No. 44;
Matches 34; Conservative 19; Mismatches 65; Indels 36; Gaps 5;

QY 113 LTTTGGKELTAFLHNGDHH--VTRIDRMPELNEAIPNDRDTPVAMATTLRLKLT 170
DB 27 LKVTAKGKSAVALIVDSRSDLDQLEVFSGEIGELFDAEKGCELNFGAGTLEVSTP 86
QY 171 GELLTLASRQOIIDMEADKVAGPLLSALPAGWFIADKSGGSGRTIATLPDGRPSR 230
DB 87 G-----VDNPL---TLPRHM-----RRNRGRVVALDDGCK 114
QY 231 --SRIVVIYTTGSGATMDERNRQIAEIGASLIKH 262
DB 115 RVARIGALNDAETHVLLIERNKRLLEVTTLELAH 148

RESULT 7
US-10-156-761-13867
Sequence 13867, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13867
LENGTH: 206
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13867

Query Match 5.1%; Score 69; DB 15; Length 206;
Best Local Similarity 25.2%; Pred. No. 53;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;

QY 70 IHYSONDLEVSPEVKEHLTDGNTV-----RELCSAATMSDNTAANLLTPTI---G 118
DB 1 VHASOG---AAGPGTESAGEEYVTMDAAGESEFEVYA-----NNSSALLKTRAVLLSG 50
QY 119 GPKR-----LTAFLHNGDHTVRLD-----RW-----EPELNEAIPNDE 152
DB 51 GDRRAEDLDNALIKADRWGRIDEPAYVRQVLYNQVSRWRLKMRRLSVAEP--PE 109
QY 153 RDTMPVAMATTLRLKLTGELLTLASRQOIIDMEADKVAGPLLSALPAGWFIADKSGGSGRTIATLPDGRPSR 205
DB 110 ASTGPDASAEELRVMRGALARTARQRTVLVLRYPEDLPEAD--VARIIGCSVGYRS 167
QY 206 IADKSGGSGRTIATLPDGRPSR---DGRPSR 232
DB 168 TTHRSRLRLRLAPELAAIGPADAEQEPSR 197

RESULT 8
US-10-101-464A-765
Sequence 765, Application US/10101464A
Publication No. US20030046728A1
GENERAL INFORMATION:
APPLICANT: Strubala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas
APPLICANT: Higgins, Colleen M.
TITLE OF INVENTION: Compositions Isolated from Plant Cells
FILE REFERENCE: 11000.1020c2
CURRENT APPLICATION NUMBER: US/10/101,464A

OY 135 TRLDREPELNEAIPNDEDTMPVAMATT 164
Db 61 TRLDREPELNEAIPNDEDTMPVAMATT 90

RESULT 2

US-10-127-816-9
; Sequence 9, Application US/10127816
; Publication No. US20030104416A1
GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.
APPLICANT: Fox, Brian A.
APPLICANT: Klucher, Kevin M.
APPLICANT: Taft, David W.

APPLICANT: Kindsvogel, Wayne R.
TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
FILE REFERENCE: 01-17
CURRENT APPLICATION NUMBER: US/10/127,816
CURRENT FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: US 60/285,408
PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 60/286,482
PRIOR FILING DATE: 2001-04-25

PRIOR APPLICATION NUMBER: US 60/341,050
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: US 60/341,105
PRIOR FILING DATE: 2001-10-22

PRIOR APPLICATION NUMBER: US 09/895,834
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/285,424
PRIOR FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 59
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 9
LENGTH: 202
TYPE: PRT

ORGANISM: Homo sapiens
US-10-127-816-9

Query Match 5.8%; Score 78; DB 15; Length 202;
Best Local Similarity 21.4%; Pred. No. 6.2;
Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEQLGRIRHYGOND--LVEYSPVTEKHL-----TDGMVYREL----- 97
Db 17 LLLAAVLTFTQADVPVPATRLPVEAKDCHIAQFSLSPRELOAFKAKDAIEKRLLEKDL 76

OY 98 -CSA-----ATMS--DNTAANLLTTTGGPRELTAFLN 129
Db 77 RCSSHLFPRAMDLOQVQERPKALQAEVALTLTKVWENMTDSALATITIGQPLHTLSHIS 136

OY 130 MGDHYT-----RLDRWEPELNEAIPNDEDTMPVAMATTLRKLTGELTL 176
Db 137 QLOTCTQLOATAPEPRSPSRRLSRWLHRLQEA-QSKETPGCLEASVTSNLFRLITDLCV 195

OY 177 ASRQOLI 183
Db 196 ANGDCV 202

RESULT 3

US-10-142-717-12
; Sequence 12, Application US/10142717
; Publication No. US20030104579A1
GENERAL INFORMATION:

APPLICANT: Immunex Corporation
APPLICANT: Baum, Peter R
APPLICANT: Mosley, Bruce A
APPLICANT: Ketchum, Randal R

APPLICANT: Taylor, Scott L
TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
; CURRENT FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 202
; TYPE: PRT

ORGANISM: Mus musculus
US-10-142-717-12

Query Match 5.8%; Score 78; DB 15; Length 202;
Best Local Similarity 21.4%; Pred. No. 6.2;
Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEQLGRIRHYGOND--LVEYSPVTEKHL-----TDGMVYREL----- 97
Db 17 LLLAAVLTFTQADVPVPATRLPVEAKDCHIAQFSLSPRELOAFKAKDAIEKRLLEKDL 76

OY 98 -CSA-----ATMS--DNTAANLLTTTGGPRELTAFLN 129
Db 77 RCSSHLFPRAMDLOQVQERPKALQAEVALTLTKVWENMTDSALATITIGQPLHTLSHIS 136

OY 130 MGDHYT-----RLDRWEPELNEAIPNDEDTMPVAMATTLRKLTGELTL 176
Db 137 QLOTCTQLOATAPEPRSPSRRLSRWLHRLQEA-QSKETPGCLEASVTSNLFRLITDLCV 195

OY 177 ASRQOLI 183
Db 196 ANGDCV 202

RESULT 4
US-10-233-926-4
; Sequence 4, Application US/10233926
; Publication No. US20030131382A1
GENERAL INFORMATION:

APPLICANT: Allen, Stephen M.
APPLICANT: Kinney, Anthony J.
APPLICANT: Falco, S. Carl

TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
FILE REFERENCE: BBI419 US NA
CURRENT APPLICATION NUMBER: US/10/233,926
CURRENT FILING DATE: 2002-09-03

PRIOR APPLICATION NUMBER: US/09/735,846
PRIOR FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Microsoft Office 97

SEQ ID NO 4
LENGTH: 149
TYPE: PRT

ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (18)

US-10-233-926-4

Query Match 5.2%; Score 69.5; DB 16; Length 149;
Best Local Similarity 20.8%; Pred. No. 30;
Matches 22; Conservative 25; Mismatches 34; Indels 25; Gaps 4;

OY 156 TMPVAMATTLRKLT-----TGELTLASROOLIDMNEADKVGPLRLSPGWTADK 209
Db 11 SLSTLSKSLUNLMAADHAAEAAPQSSQEEEDKREAGCGGDV-----EVADR 61

OY 210 SGAGERSRGITIALGPDGKPSRIYV-----IYTGQQAFTDERNR 250
Db 62 GGGCGAANGGI-----PEGRIYVADGIYDLFHFHGHAKSLDQAKR 102

RESULT 5

US-10-156-761-8136
; Sequence 8136, Application US/10156761
; Publication No. US20030119018A1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds
(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_30E_37S
Sequence: 1 HPETLVKVKDAEDQLGARVQ.....TMDERNKQIAETIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppa/PCFUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|-------------------|
| 1 | 463 | 34.4 | 94 | US-10-102-806-760 | Sequence 760, App |
| 2 | 78 | 5.8 | 202 | US-10-127-816-9 | Sequence 9, Appl |
| 3 | 78 | 5.8 | 202 | US-10-142-717-12 | Sequence 12, Appl |
| 4 | 69.5 | 5.2 | 149 | US-10-233-926-4 | Sequence 4, Appl |
| 5 | 69 | 5.1 | 149 | US-10-156-761-8136 | Sequence 8136, Ap |
| 6 | 69 | 5.1 | 178 | US-09-738-626-5680 | Sequence 5680, Ap |
| 7 | 69 | 5.1 | 208 | US-10-156-761-13867 | Sequence 13867, A |
| 8 | 67.5 | 5.0 | 184 | US-10-101-464A-765 | Sequence 765, App |
| 9 | 67.5 | 5.0 | 187 | US-09-948-018-8 | Sequence 8623, Ap |
| 10 | 67 | 5.0 | 194 | US-10-156-761-8623 | Sequence 8623, Ap |
| 11 | 67 | 5.0 | 202 | US-10-127-816-11 | Sequence 11, Appl |
| 12 | 65.5 | 4.9 | 191 | US-09-764-870-390 | Sequence 390, App |
| 13 | 65.5 | 4.9 | 191 | US-10-125-540-390 | Sequence 390, App |
| 14 | 65 | 4.8 | 179 | US-09-764-868-757 | Sequence 757, App |
| 15 | 65 | 4.8 | 179 | US-10-106-658-4858 | Sequence 4858, Ap |

| | | | | | |
|----|------|-----|-----|---------------------|--------------------|
| 16 | 65 | 4.8 | 202 | US-10-189-346-12 | Sequence 12, Appl |
| 17 | 64.5 | 4.8 | 189 | US-09-815-242-5862 | Sequence 5862, App |
| 18 | 64.5 | 4.8 | 189 | US-09-815-242-12979 | Sequence 12979, A |
| 19 | 64.5 | 4.8 | 191 | US-10-156-761-8978 | Sequence 8978, Ap |
| 20 | 64.5 | 4.8 | 193 | US-09-805-354-8 | Sequence 8, Appl |
| 21 | 64.5 | 4.8 | 193 | US-10-144-259-21 | Sequence 8, Appl |
| 22 | 63.5 | 4.7 | 202 | US-10-156-761-13460 | Sequence 13460, A |
| 23 | 63.5 | 4.7 | 206 | US-09-738-626-5425 | Sequence 5425, Ap |
| 24 | 62.5 | 4.6 | 144 | US-10-131-406-4 | Sequence 4, Appl |
| 25 | 62.5 | 4.6 | 100 | US-09-882-227-230 | Sequence 230, App |
| 26 | 62.5 | 4.6 | 130 | US-10-156-761-9507 | Sequence 9507, Ap |
| 27 | 62 | 4.6 | 162 | US-09-738-626-4796 | Sequence 4796, Ap |
| 28 | 62 | 4.6 | 174 | US-09-864-761-35777 | Sequence 35777, A |
| 29 | 62 | 4.6 | 185 | US-10-101-664A-616 | Sequence 616, App |
| 30 | 62 | 4.6 | 190 | US-09-738-626-4637 | Sequence 4637, Ap |
| 31 | 61.5 | 4.6 | 88 | US-10-100-252-6 | Sequence 6, Appl |
| 32 | 61.5 | 4.6 | 142 | US-09-862-027-31 | Sequence 31, Appl |
| 33 | 61.5 | 4.6 | 187 | US-10-156-761-12111 | Sequence 12111, A |
| 34 | 61.5 | 4.6 | 195 | US-09-738-626-6624 | Sequence 6624, Ap |
| 35 | 61.5 | 4.6 | 202 | US-10-189-346-16 | Sequence 16, Appl |
| 36 | 61 | 4.5 | 70 | US-09-864-761-41555 | Sequence 41555, A |
| 37 | 61 | 4.5 | 177 | US-09-791-932-69 | Sequence 69, Appl |
| 38 | 61 | 4.5 | 189 | US-09-764-870-523 | Sequence 523, App |
| 39 | 61 | 4.5 | 189 | US-10-125-540-523 | Sequence 523, App |
| 40 | 61 | 4.5 | 203 | US-09-800-729-154 | Sequence 154, App |
| 41 | 60.5 | 4.5 | 152 | US-10-156-761-14346 | Sequence 14346, A |
| 42 | 60.5 | 4.5 | 196 | US-10-219-220-265 | Sequence 265, App |
| 43 | 60 | 4.5 | 134 | US-09-768-235B-40 | Sequence 40, Appl |
| 44 | 60 | 4.5 | 184 | US-09-798-029-4 | Sequence 4, Appl |
| 45 | 60 | 4.5 | 184 | US-09-795-926-46 | Sequence 46, Appl |

ALIGNMENTS

RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760

Query Match 34.4%; Score 463; DB 15; Length 94;

Best Local Similarity 98.9%; Pred. No. 2.9e-40;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 NDLYEYSPYTEKILTDGNTVRELCSAIIYTSNDTANLLTTIGGPKELTATLHNGCHV 134
|||||
DB 1 NDLYEYSPYTEKILTDGNTVRELCSAIIYTSNDTANLLTTIGGPKELTATLHNGCHV 60

NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 24921
 LENGTH: 203
 TYPE: PRT
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-24921

Query Match
 Best Local Similarity 27.1%; Score 64.5; DB 4; Length 203;
 Matches 35; Conservative 16; Mismatches 41; Indels 37; Gaps 7;

QY 135 TLDNPEPL-----NEAIPNDERDTMPVAMATTLRLKLTGELLTLASROQLIDME 187
 DB 34 TPLDSESERLRVSLAWRKAMPMPMAQARPMKLAEL--RALTGSLTILSR-----DMS- 86
 QY 188 ADKVAQPLRSALPACGFITDKSGAGERSGCTTALALPGDKPSRIIVITTTGSGQATME 247
 DB 87 -----TSNAPPGGDI-----SPGWLASTRLGK-AKLAIVATY---AALKV 123
 QY 248 RNRQIAEIG 256
 DB 124 PRRYSSTIG 132

RESULT 14
 US-09-239-909-2
 Sequence 2, Application US/09239909
 Patent No. 6284952
 GENERAL INFORMATION:
 APPLICANT: Kumho Petrochemical Co. Ltd.
 TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve N
 FILE REFERENCE: P99-2-6
 CURRENT FILING DATE: 1999-01-29
 EARLIER APPLICATION NUMBER: EP 99300136.1
 EARLIER FILING DATE: 1999-01-08
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: KOPATIN 1.0
 SEQ ID NO 2
 LENGTH: 150
 TYPE: PRT
 ORGANISM: G. max calmodulin4 (SCAM4)
 US-09-239-909-2

Query Match
 Best Local Similarity 27.9%; Score 64; DB 3; Length 150;
 Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

QY 73 SONDLVEYSPYTEKHLTDG---MTVELCSAATMSDNTANLLITF-----GGPRE 122
 DB 6 SEQIVDRKRAFGLEDKDGCGTVELATVIRSLDQPTREELQDMISEVDADNGTIE 65
 QY 123 LTAFLNMGDHYTRLDRWEPELNEAIPNDERDTMPVAMATTLRLKLT--TGELLTLASRO 180
 DB 66 FDEFSLAKKAKDDDA--EELKEAFKVFCKDNGYIS-ASELKHVMINLGEKLTDEEVE 123
 QY 181 QLIDMEAD 189
 DB 124 QMI--KQAD 130

RESULT 15
 US-08-624-677A-2
 Sequence 2, Application US/08624677A
 Patent No. 6476192
 GENERAL INFORMATION:
 APPLICANT: Lally, Nicola C.
 APPLICANT: Jenkins, Mark C.
 APPLICANT: Dubey, Jitender P.
 TITLE OF INVENTION: Antigens useful for the serodiagnosis of
 TITLE OF INVENTION: Neosporosis
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Graeter, Janelle S.
 STREET: Rm. 411, Bldg. 005, BARC-W
 CITY: Beltsville
 STATE: MD
 COUNTRY: USA
 ZIP: 20705
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/624,677A
 FILING DATE: 15-APR-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Graeter, Janelle S.
 REGISTRATION NUMBER: 35,024
 REFERENCE/DOCKET NUMBER: 0228.95
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 301-504-6629
 TELEFAX: 301-504-5060
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 203 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-624-677A-2

Query Match
 Best Local Similarity 23.8%; Score 64; DB 4; Length 203;
 Matches 39; Conservative 23; Mismatches 72; Indels 30; Gaps 7;
 QY 54 AVLSRIDAGQEQGRRIHYSQNDLVEYSPYTEKHLTDGMTVELCSAA--ITMSDNTAN 111
 DB 33 AGVSNYDGDGDDAAGNVD-----SDVTDAITGDEMPRVVSGQPHRTQKSLRK 83
 QY 112 LLLTTIGPKRELFALNMGDHY-TRLDRWEPELNEAIPNDER-DTTPVA----- 160
 DB 84 LAVPVGA---LTSYL--VADRYLPELTSAGESTESIPGKRKRTAGAGIALVAANAFA 138
 QY 161 ---MATTLRKLTGELLTLASROQLIDMEADKVGPLNSALP 201
 DB 139 GIGLARTFRHFVPRKSKYVASEDSALGNSSEQYVEGTVNGSSDP 182

Search completed: September 10, 2003, 12:31:04
 Job time : 12.8571 secs

Best Local Similarity 21.6%; Pred. No. 29;
Matches 46; Conservative 32; Mismatches 72; Indels 63; Gaps 8;

OY 15 LGARVGYIELDNSELSEFRSEEPFPMSTFKVLCGAVLSRIDAGQ--EQGRRIHY 72
DB 1 LKRNIGLVLDISTIELSPICRKSVSRPK-----AAAFVDAISFIOELPORD- 54
OY 73 SONDVEYSPYTERKHLTDGMYRELCSAATMSDNTAANLLTTIGPKELTAFLHMGD 132
DB 55 -----SPYSEKSSSTGTGROLLAFARTVASO-----PKLTI----- 86
OY 133 HVTLRDREPELNEIPDERDTPMPVAMATLRLKLTGELLTLASROQLDMWEADKVA 192
DB 87 -----LDEATANDSETESLV--QASLAKRROGR-TTATNHRISTODANCI- 131
OY 193 GPLRSALPAGMFTADKSGAGRGSRGIIAAG 225
DB 132 -----YVLDKRIIESGTHEELLAIG 152

RESULT 8

US-09-252-991A-29942
Sequence 29942, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29942
LENGTH: 178
TYPE: PRN
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (116)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match
Best Local Similarity 26.8%; Score 65.5; DB 4; Length 178;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

OY 110 ANLLTTIG--GPKELTA-----LHMGDHYTLDRKEPLNEAIPDERDTPMPVMA 162
DB 25 ANVLVELLAQQRFGELAAHAGLAHHPGSHVAR-----AGEDLP--RDVDLGAAA 74
OY 163 TTKRLKLLGELL-----TLASROQLDMWEADKVAGPLRLSALPAGMFTADSG 211
DB 75 LHHHLYLAGALOGHEHGGIDGTAGSEAVVG--QOEVAQPVY--GLQAGLFLMVGD 130
OY 212 AGERSGRIIAGLPDGK 229
DB 131 A-----LVGVGQAGK 141

RESULT 9

US-08-557-122A-6
Sequence 6, Application US/08557122A
Patent No. 5879664

GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5879664 No. 5879664disk of No. 5879664th America, Inc.

STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-122A-6

Query Match
Best Local Similarity 4.8%; Score 65; DB 2; Length 174;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

OY 7 KYKADMDLGAR-VGTYELDNLSEFRSEEPFPMSTFKVLCGAVLSRIDAGQEQ 65
DB 67 KTEQATELKEKNIPYKVKDCTEEALCRDQGEVGYPTLKFRL-----DAVKPY 117
OY 66 LGRR-----IHS-QNDLVEYSPYTERKHLTDGMYRELCSAATMSDNTAANLLTT 116
DB 118 QGARTFAIVSYMKOSLPVSPVYEPENLEIKTKIVIGYIASDQYANDIFTT 174

RESULT 10

US-09-262-666-6
Sequence 6, Application US/09262666
Patent No. 6346244

GENERAL INFORMATION:

APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6346244 No. 6346244disk of No. 6346244th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/262,666
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/557,122

FILING DATE: 11-DEC-1995

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980.204-US

```

      :: : | | | | : | | |
111  SMAPLAAVSTWSPVITTAIPAAWHSATATASSRG 147
Db

```

RESULT 5
US-09-413-814-87
Sequence 87. Application US/09413814

```

; Patent No. 6225064
;
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoffe, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
;
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE OF INVENTION: heteropolypeptide compounds
; FILE REFERENCE: PCT/US 99/23535
;
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 67
; LENGTH: 198
;
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
;
; US-09-413-814-87

```

| | | | | | | | |
|-----------------------|--------|--------------|---------|------------|-----|--------|------|
| Query Match | 4.9%; | Score | 66.5; | DB | 3; | Length | 198; |
| Best Local Similarity | 22.4%; | Pred. | No. 28; | | | | |
| Matches | 62; | Conservative | 31; | Mismatches | 77; | Indels | 107; |
| | | | | | | Gaps | 15; |

```

QY      1  HPELLVKKRDMEDQDGLARGAYIELDNGSGIIIEEFSEERPEPMSTF-----KULLGAV  55
D      7  NPEAVDAVKDKTS--AAFG-----DATVHKVLBEIGAQVETALVTFGHEPSPVLCVAS  60
QY      56  LSRIDAGEQLCRRIHYSONDIVEGSPYTEKHLTDGMYRELCSAATMSDNTAANLLT  115
D      61  LVR-----MGVRIIARA-----TDSQAD--TLR  82
QY      116  TIGCPKELTAFILHNMGDHYTRLDRWEPLEALIPNDERDTPVPMATFLRKLITGELLT  175
D      83  AVG-----ATRVQLLETENGRV-----GADIMPLAQ-----DLUD  114
QY      176  LASROOLDMMENDKVGAPLRLSALPAGMTIADK---SGAGER-----GSKGIIAALG  225
D      115  LASHRYVVPW---NAHGPLVQOTL--GSKLRQRYRINVLGVGRPHTKRKPRGDKRLEAPT  165
QY      226  PDGKPSRIVIVITTSQATM-----DERNQIQAIG  256
D      170  PD-----TYIRGDITLLNGDSDDVSFFVAEFG  197

```

```

RESULT 6
US-09-010-809-19
Sequence 19, Application US/09010809B
Patent No. 6090601
GENERAL INFORMATION:
APPLICANT: Gustafsson, Claes
APPLICANT: Bellach, Mary C.
TITLE OF INVENTION: Epithallone Polyketide Synthases and Encoding DNA
FILE REFERENCE: 30062-20020.00
CURRENT APPLICATION NUMBER: US/09/010,809B
CURRENT FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 23

```

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 158
; TYPE: PRN
; ORGANISM: Sorangium cellulosum
; US-09-010-809-19

```

| | | | | |
|-----------------------|-------|---------------|-------|----------------|
| Query Match | 4.9% | Score 65.5; | DB 3; | length 158; |
| Best Local Similarity | 27.6% | Pred. No. 26; | | |
| Matches | 37; | Conservative | 18; | Mismatches 48; |
| | | | | Indels 31; |
| | | | | Gaps 6; |

[illegible]

RESULT 7
US-08-858-207A-519
; Sequence 519, Application US/08858207A

1 GENERAL INFORMATION:
2 APPLICANT: Black, Michael
3 APPLICANT: Hodgson, John
4 APPLICANT: Knowles, David
5 APPLICANT: Nichols, Richard
6 APPLICANT: Stodols, Robert
7 TITLE OF INVENTION: No. 634632bel Compounds
8 NUMBER OF SEQUENCES: 552
9 CORRESPONDENCE ADDRESS:
10

1 NEED: /09 Sweeland road
 2 CITY: King of Prussia
 3 STATE: PA
 4 COUNTRY: USA
 5 ZIP: 19406-0939
 6
 7 COMPUTER READABLE FORM:
 8
 9 MEDIUM TYPE: Diskette
 10
 11 COMPUTER: IBM Compatible
 12
 13 OPERATING SYSTEM: DOS
 14
 15 SOFTWARE: Fastset for Windows Version 2.0
 16
 17 CURRENT APPLICATION DATA:
 18
 19 APPLICATION NUMBER: 05/08/858,207A
 20
 21 FILING DATE: 09-MAY-1997
 22

```

; INFORMATION FOR SEQ ID NO: 519:
; SEQUENCE CHARACTERISTICS:
;

```

```

; LENGTH: 170 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 63483286
;
US-08-658-207A-519

```

Query Match 4.9%; Score 65.5; DB 4; Length 170;

```
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 202:
SEQUENCE CHARACTERISTICS:
LENGTH: 124 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-202

Query Match
Best Local Similarity 5.4%; Score 72.5; DB 4; Length 124;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;

QY 130 MGDHV---TRIDRREPELN---EALPNDERDT---TMPVAMATTLKLLTGELTLASRQ 180
DB 12 MGDGIGMERGCRGTGNTQCPLRYVPGDSEPTIDGRASPEDTLIT---NLISPTLMSPPS 69
QY 181 QLDIWEA-DKVAAGPLL-----RSALPAGWFIADKSGSGSGGIIIALGPDGKPSR 232
DB 70 RDDDWVEPFDALGTAFAVDATGDKATPAVAGT-----GASTGCS-GIIASLSPPRQPAR 123

RESULT 3
US-08-991-890-4
Sequence 4, Application US/08991890
Patent No. 6114307
GENERAL INFORMATION:
APPLICANT: Jaspers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
STIMULATING PANCREATIC ISLET CELL REGENERATION
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:
```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-991-890-4

Query Match
Best Local Similarity 5.3%; Score 71; DB 3; Length 159;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

QY 49 VLLCGAVLSRIDAGE-----OLGRRIHYSN-----DLVEYSPTERKH 87
DB 12 LLLCGAVF--VSPQELHAEFGCRHHHHHGGSGAELNCGPFGKHLISYCPMEPKT 69
QY 88 LTDGTVRELCSAATMSDNTAANLLTTIGGPRELTAFLNMGDHYTRLDWRPELNEA 147
DB 70 FT-----TTGCGWLES-GREKEMVSTSNKGOAL-----GTISEF 105
QY 148 IPNDERDTMPVAMA-TTLRKLL 169
DB 106 IPNLSPFLKKPLSECPFLKII 128

RESULT 4
US-09-252-991A-24241
Sequence 24241, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24241
LENGTH: 197
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match
Best Local Similarity 5.2%; Score 70.5; DB 4; Length 197;
Matches 37; Conservative 22; Mismatches 43; Indels 55; Gaps 7;

QY 94 VRELCSAATMSDNTAANLLTTIGGPRELTAFLNMGDHYTRLDWRPELNEAIPNDER 153
DB 12 VARTSSARTSASASLASVPLPIATP---TSARFSAGASLT-----PSPVI 55
QY 154 DFTMPVA-MATTLRKLLTGELTLA-----SQQLIDWEADKVAQP----- 194
DB 56 ATTSPLACRACRCTRSLCSG--LARAANTSTPGSTRRR--WSSNSISAPVAGRSMPIP 110
QY 195 -----ILRSALPAGWFIADKSGAGERG 216
```

Thu Sep. 11 16:08:01 2003

seq2_30e_37s.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using BW model

Run on: September 10, 2003, 12:19:51 : Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_30E_37S
Perfect score: 1345
Sequence: 1 HPEVLVKNDAEDQAGAVG.....TMDERNQIAETGASLIRHW 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Dackfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|------------------------|--------------------|
| 1 | 74 | 5.5 | 169 | 4 US-09-328-352-4491 | Sequence 4491, App |
| 2 | 72.5 | 5.4 | 124 | 4 US-08-311-731A-202 | Sequence 202, App |
| 3 | 71 | 5.3 | 159 | 3 US-08-991-890-4 | Sequence 4, Appl1 |
| 4 | 70.5 | 5.2 | 197 | 4 US-09-252-991A-24241 | Sequence 24241, A |
| 5 | 65.5 | 4.9 | 198 | 3 US-09-413-814-87 | Sequence 87, Appl1 |
| 6 | 65.5 | 4.9 | 158 | 3 US-09-010-809-19 | Sequence 19, Appl1 |
| 7 | 65.5 | 4.9 | 170 | 4 US-08-858-207A-519 | Sequence 519, App |
| 8 | 65.5 | 4.9 | 178 | 4 US-09-252-991A-29942 | Sequence 29942, A |
| 9 | 65 | 4.8 | 174 | 4 US-08-557-122A-6 | Sequence 6, Appl1 |
| 10 | 65 | 4.8 | 174 | 4 US-09-262-666-6 | Sequence 12, Appl1 |
| 11 | 65 | 4.8 | 200 | 2 US-08-557-122A-12 | Sequence 12, Appl1 |
| 12 | 65 | 4.8 | 200 | 4 US-09-262-666-12 | Sequence 24921, A |
| 13 | 64.5 | 4.8 | 203 | 4 US-09-252-991A-24921 | Sequence 2, Appl1 |
| 14 | 64 | 4.8 | 150 | 3 US-09-239-909-2 | Sequence 2, Appl1 |
| 15 | 64 | 4.8 | 203 | 4 US-08-624-677A-2 | Sequence 4, Appl1 |
| 16 | 62.5 | 4.6 | 144 | 2 US-08-225-480-4 | Sequence 2, Appl1 |
| 17 | 62.5 | 4.6 | 144 | 4 US-09-118-445-4 | Sequence 22681, A |
| 18 | 62 | 4.6 | 132 | 4 US-09-252-991A-22681 | Sequence 17611, A |
| 19 | 62 | 4.6 | 204 | 4 US-09-252-991A-17611 | Sequence 31, Appl1 |
| 20 | 61.5 | 4.6 | 142 | 4 US-09-345-473E-31 | Sequence 30363, A |
| 21 | 61 | 4.5 | 169 | 2 US-08-680-849-2 | Sequence 2, Appl1 |
| 22 | 60.5 | 4.5 | 167 | 3 US-09-004-053-2 | Sequence 25578, A |
| 23 | 60.5 | 4.5 | 144 | 4 US-09-252-991A-25578 | Sequence 4218, App |
| 24 | 60 | 4.5 | 158 | 4 US-09-107-532A-4218 | Sequence 195, App |
| 25 | 60 | 4.5 | 181 | 2 US-08-482-142-195 | Sequence 195, App |
| 26 | 60 | 4.5 | 181 | 2 US-08-478-572-195 | Sequence 195, App |
| 27 | 60 | 4.5 | 181 | 2 US-08-478-572-195 | Sequence 195, App |

| | | | | | |
|----|------|-----|-----|------------------------|--------------------|
| 28 | 60 | 4.5 | 181 | 3 US-08-484-296-195 | Sequence 195, App |
| 29 | 60 | 4.5 | 184 | 4 US-09-795-926-46 | Sequence 46, Appl1 |
| 30 | 59.5 | 4.4 | 103 | 4 US-09-732-210-1282 | Sequence 1282, App |
| 31 | 59.5 | 4.4 | 146 | 4 US-09-134-001C-5269 | Sequence 5269, App |
| 32 | 59.5 | 4.4 | 160 | 4 US-09-252-991A-24737 | Sequence 24737, A |
| 33 | 59.5 | 4.4 | 168 | 2 US-08-598-873-3 | Sequence 3, Appl1 |
| 34 | 59.5 | 4.4 | 168 | 3 US-08-605-430-3 | Sequence 22351, A |
| 35 | 59.5 | 4.4 | 171 | 4 US-09-252-991A-22351 | Sequence 30622, A |
| 36 | 59 | 4.4 | 136 | 4 US-09-252-991A-30622 | Sequence 3640, App |
| 37 | 59 | 4.4 | 148 | 4 US-09-134-001C-3640 | Sequence 3640, App |
| 38 | 59 | 4.4 | 171 | 4 US-09-107-532A-3979 | Sequence 3979, App |
| 39 | 59 | 4.4 | 180 | 4 US-09-194-905-5 | Sequence 48, Appl1 |
| 40 | 59 | 4.4 | 189 | 4 US-08-671-548C-48 | Sequence 47, Appl1 |
| 41 | 58.5 | 4.3 | 102 | 3 US-08-479-744A-47 | Sequence 281, App |
| 42 | 58.5 | 4.3 | 102 | 3 US-08-280-757B-47 | Sequence 412, App |
| 43 | 58.5 | 4.3 | 174 | 4 US-09-072-596-281 | Sequence 412, App |
| 44 | 58 | 4.3 | 72 | 4 US-08-858-207A-412 | Sequence 6, Appl1 |
| 45 | 58 | 4.3 | 168 | 1 US-08-451-947-6 | Sequence 6, Appl1 |

ALIGNMENTS

```
RESULT 1
US-09-328-352-4491
Sequence 4491, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328, 352
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4491
LENGTH: 169
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-4491
```

Query Match 5.5%: Score 74; DB 4; Length 169;
Best Local Similarity 23.3%; Pred. No. 3.3;
Matches 37; Conservat 34; Mismatches 62; Indels 26; Gaps 7;

```
OY 118 GGPRLTAPFLNMGDHYRLDRWPELEAIPN-DEDTTPVAMATTLKLTG-----E 172
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 GGENTIMKMLAQON---RFDQWEOIYOYDLRLVYRERIVVFTTIFVYVIVGYSLMK 61
OY 173 LPTLASRQO-----LIDWMDKRVAGPLRSALPGWFTADKSGAGER--GSRGIIA 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 MHSIAEQOQKRLNDLKMVMWQSNV-----TMKPAWLELDKSGKIQVAAQOQGLTV 115
OY 223 ALGPDKPSRIVYITTSQATMDERNQIAETGASLIR 261
DB 116 SSQNGEQLQIV--THQNYATLANFLTQLAQWLSIQK 152

RESULT 2
US-08-311-731A-202
Sequence 202, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
DIAGNOSTICS AND THERAPEUTICS
TITLE OF INVENTION:
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSER: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
```


THIS PAGE BLANK (USPTO)


```

RESULT 14
AAW55524
ID AAW55524 standard; Protein; 200 AA.
AC AAW55524;
XX
XX
XX 02-JUL-1998 (first entry)
XX
XX
XX H. pylori ORF 29ep10720_24432762_c3_39 cellular protein.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
XX Identification; binding compound; bacteria; life cycle; activator;
XX Inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX Helicobacter pylori.
XX
XX MO9737044-A1.
XX
XX 09-OCT-1997.
XX
XX 27-MAR-1997; 97WO-US05223.
XX
XX 06-DEC-1996; 96US-0761318.
XX 29-MAR-1996; 96US-0625811.
XX 02-APR-1996; 96US-0758731.
XX 25-OCT-1996; 96US-0736905.
XX 28-OCT-1996; 96US-0738859.
XX
XX (ASTRA ) ASTRA AB.
XX
XX Alm RA, Smith D;
XX
XX WPI; 1997-503122/46.
XX N-PSDB; AAV24933.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
XX polypeptide(s) - useful in vaccines to treat or prevent H. pylori
XX infection and for diagnosis of H. pylori infection
XX
XX Claims 14,93; Page 726-727; 1145pp; English.
XX
XX This sequence is a H. pylori cellular protein.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds,
XX useful as potential H. pylori life cycle activators or inhibitors. The
XX DNA and probes derived from it may be used for the identification of
XX H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
XX acid sequences complementary to the DNA act as antisense sequences and
XX can be used to prevent the translation of H. pylori mRNA. Antibodies
XX against the protein can be used in immunoassays to evaluate the abundance
XX and distribution of H. pylori-specific antigens. The genomic sequence of
XX H. pylori (ATCC 55679) was determined from overlapping contigs generated
XX by mechanically shearing the bacterial DNA. The sequences were analysed
XX for ORF of at least 180 nucleotides, and the predicted coding regions
XX vaccine development, the amino acid sequences predicted from various ORF
XX were analysed for significant homology to other known or exported
XX membrane proteins. Having identified and determined the sequences of
XX interest, particular regions can be isolated from H. pylori by PCR
XX amplification for recombinant polypeptide production, e.g. in E. coli
XX hosts.
XX
XX Sequence 200 AA;
XX
Query Match 6.0%; Score 81; DB 18; Length 200;
Best Local Similarity 26.0%; Pred. No. 5.4;
Matches 51; Conservative 32; Mismatches 43; Indels 70; Gaps 13;
OY 22 IELDUNSGELIESF---RSEERPPMSTFVYLGCAYLSRIDAGOEOLGRRIHYSQNDIV 78
DB 11 VILDNDSDKSMERFAIRAKERPTPSLF-----NRSRGFSDTLKOWMSKEYENIILI 61
OY 79 ----EYSPYTERKHLTDGMVREICSAITMSDNTANULL--TTIGPKEL-TATLHNMG 131

```

```

DB 62 DTKGEYSKEIOT-----AMLS-----NIVLVPT-----PSQDDREVLNML 99
OY 132 DHVTRIDKWEPELNE--AIPNDEKDTMPVAMATTLRLKLGELTLASROOLIDWME- 187
DB 100 ERIQQLQ-----ELNENLRALVYNRPPTIP-----TLKRGQALIERIKE 139
OY 188 ---ADKVAGPLRSAL 200
DB 140 NNPEDRIT--LLESST 153

RESULT 15
AAU45138
ID AAU45138 standard; Protein; 158 AA.
AC AAU45138;
XX
XX 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #6034.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-206841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIAX CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX N-PSDB; AAS59525.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
XX treating acne vulgaris -
XX
XX Example 1; SEQ ID NO 6333; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

```


DB 140 VVGTYIG 145

RESULT 10
ABG27935

ID ABG27935 standard; Protein; 101 AA.

XX ABG27935;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #27926.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI: 2001-639362/73.

DR N-PSDB; AAS92122.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID No 58294; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 101 AA;

Query Match 27.7%; Score 372; DB 22; Length 101;

Best Local Similarity 100.0%; Pred. No. 1e-30;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 DTTMPVAMATLTKLTGELLTASROQLDMMADRVAGPLLSALPAGWTIAKSGAG 213

DB 10 DTTMPVAMATLTKLTGELLTASROQLDMMADRVAGPLLSALPAGWTIAKSGAG 69

QY 214 ERGSRGIIALGPD 227

DB 70 ERGSRGIIALGPD 83

RESULT 11
ABG27919

ID ABG27919 standard; Protein; 88 AA.

XX ABG27919;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #27910.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI: 2001-639362/73.

DR N-PSDB; AAS92106.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID No 58278; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridization probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensic, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 88 AA;

Query Match 24.2%; Score 325; DB 22; Length 88;

Best Local Similarity 80.0%; Pred. No. 6.0e-26;

Matches 68; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

QY 4 TLVKVDKEDQAGRGYIELDINSGEILSPSEERFPMSTFKVLLCGAVLSRIDAGQ 63

DB TLVKVDKEDQAGRGYIELDINSGEILSPSEERFPMSTFKVLLCGAVLSRIDAGQ 63

```

XX AAB59052;
AC 27-MAR-2001 (first entry)
XX
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX
KW Human: breast cancer; cytostatic; immunosuppressive;
KW neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;
KW antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;
KW Addison's disease; allergy; autoimmune hemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.
XX
OS Homo sapiens.
XX
PN M020005173-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-611515/58.
XX
DR N-PSDB; AAF21955.
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 11; Page 1228; 1299pp; English.
XX
CC Sequences AAB591614 - AAB592031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAB59129 - AAB592031 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist
CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neutropenic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;
CC antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and antagonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune hemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 94 AA;

```

Query Match 34.4%; Score 463; DB 21; Length 94;
 Best Local Similarity 98.9%; Pred. No. 3.2e-40;
 Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 75 NDLEYSPTTEKHLTDGKMTVRELCSAITSMDNTAANLLTTTGGPKELTFLHNGDHY 134
DB 1 NDLEYSPTTEKHLTDGKMTVRELCSAITSMDNTAANLLTTTGGPKELTFLHNGDHY 60
OY 135 TRLDWEPELNEAIPDERDRTTPVAMATT 164
DB 61 TRLDWEPELNEAIPDERDRTTPVAMATT 90

```

```

RESULT 9
AAB60628
ID AAB60628 standard; Protein; 159 AA.
XX
AC AAB60628;
XX
DE 25-MAR-2003 (updated)
DE 17-JUN-1991 (first entry)
XX
KW Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
KW Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
KW fusion protein; beta-lactamase.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FH Protein 1..96
FT /label= beta-lactamase
FT Misc-difference 97..100
FT /label= adaptor
FT Protein 101..159
FT /label= beta-urogastrone
XX
PN DE3523634-A.
XX
PD 09-JAN-1986.
XX
PF 02-JUL-1985; B5DE-3523634.
XX
PR 02-JUL-1984; 84JP-0137691.
XX
PA (EART ) EARTH CHEM CO LTD.
XX
PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;
PI Matsushiro S;
XX
DR WPI; 1986-015031/03.
XX
DR N-PSDB; AAN60632.
XX
PT New gene for expression of beta-urogastrone - its deriva., plasmid(s)
PT and transformed cells contg. it.
XX
PS Disclosure; Page 59-61; 92pp; German.
XX
CC The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 159 AA;

```

Query Match 29.4%; Score 396; DB 7; Length 159;
 Best Local Similarity 67.5%; Pred. No. 6.3e-33;
 Matches 85; Conservative 9; Mismatches 20; Indels 12; Gaps 2;

```

OY 1 HPEITVKKYKDAEDDQGARVGYIELDNGSILSFREERFPNMSTFVLLCGAVLSRD 60
DB 24 HPEITVKKYKDAEDDQGARVGYIELDNGSILSFREERFPNMSTFVLLCGAVLSRD 83
OY 61 AGOEOLGRIRHYSNDLVEYS-----PYTEKHLTDGKMTVRELCSAITSMDNTAANL 112
DB 84 AGOEOLGRIRHYSNDLVEYSARNSDECPESH---DGYCLHDGVCYKYTEALDKYACNC 139
OY 113 LTTTIG 118
DB :: ::

```

```
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251979.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-483426/52.
XX
DR N-PSDB; AAK63652.
XX
PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Claim 11; SEQ ID NO 18464; 3071bp + Sequence Listing; English.
XX
CC AAK64931 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK64942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 129 AA;
XX
Query Match 34.9%; Score 469; DB 22; Length 129;
Best Local Similarity 95.8%; Pred. No. 1.2e-40;
Matches 92; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 HPELVKVKDAEDQAGARVGYIELDINSEIFSESEFPMPKSTFKVLLCGAVLSRID 60
DB 28 HPELVKVKDAEDQAGARVGYIELDINSEIFSESEFPMPKSTFKVLLCGAVLSRID 87
QY 61 AGQEGLRRIHYSQNDLVEYSPVTEKHLTDGMTVRE 96
DB 88 AGQEGLRRIHYSQNDLVEYSPVTEKHLTDGMTXXE 123
XX
RESULT 7
ABG27917
ID ABG27917 standard; Protein; 182 AA.
XX
AC ABG27917;
```

```
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27908.
XX
KW Human; Chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001MO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
XX
DR N-PSDB; AAS92104.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 58276; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 182 AA;
XX
Query Match 34.9%; Score 469; DB 22; Length 182;
Best Local Similarity 86.4%; Pred. No. 1.9e-40;
Matches 95; Conservative 4; Mismatches 7; Indels 4; Gaps 1;
QY 1 HPELVKVKDAEDQAGARVGYIELDINSEIFSESEFPMPKSTFKVLLCGAVLSRID 60
DB 24 HPELVKVKDAEDQAGARVGYIELDINSEIFSESEFPMPKSTFKVLLCGAVLSRID 83
QY 61 AGQEGLRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCGSAITMSDNTAA 110
DB 84 AGQEGLRRIHYSQNDLVEYSPVTEKHLTDGMTER-----LENSFDNTAS 129
XX
RESULT 8
AAB59052
ID AAB59052 standard; Protein; 94 AA.
```

OY 1 HPEETLVKVADEQLGARVGIIEIDNGSELLESFRSPERFPMNSPFYLLCGAVLSRID 60
DB 28 HPEETLVKVADEQLGARVGIIEIDNGSELLESFRSPERFPMNSPFYLLCGAVLSRID 87
OY 61 AGQEOIGRRIRHYSNDLVVEYSPVTEKHILTDGNTVRE 96
DB 88 AGQEOIGRRIRHYSNDLVVEYSPVTEKHILTDGNTVRE 123

RESULT 6
AAM90871
ID AAM90871 standard; Protein: 129 AA.
XX AAM90871;
AC
XX
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen SEQ ID NO:18464.
XX
KW Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN W0200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184654.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225277.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 16-AUG-2000; 2000US-0225779.
PR 22-AUG-2000; 2000US-022681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.

| | | |
|----|-------------|----------------|
| PR | 08-NOV-2000 | 2000US-0246613 |
| PR | 17-NOV-2000 | 2000US-0249207 |
| PR | 17-NOV-2000 | 2000US-0249208 |
| PR | 17-NOV-2000 | 2000US-0249209 |
| PR | 17-NOV-2000 | 2000US-0249210 |
| PR | 17-NOV-2000 | 2000US-0249211 |
| PR | 17-NOV-2000 | 2000US-0249212 |
| PR | 17-NOV-2000 | 2000US-0249213 |
| PR | 17-NOV-2000 | 2000US-0249214 |
| PR | 17-NOV-2000 | 2000US-0249215 |
| PR | 17-NOV-2000 | 2000US-0249216 |
| PR | 17-NOV-2000 | 2000US-0249217 |
| PR | 17-NOV-2000 | 2000US-0249218 |
| PR | 17-NOV-2000 | 2000US-0249244 |
| PR | 17-NOV-2000 | 2000US-0249245 |
| PR | 17-NOV-2000 | 2000US-0249264 |
| PR | 17-NOV-2000 | 2000US-0249265 |
| PR | 17-NOV-2000 | 2000US-0249297 |
| PR | 17-NOV-2000 | 2000US-0249299 |
| PR | 17-NOV-2000 | 2000US-0249300 |
| PR | 01-DEC-2000 | 2000US-0250160 |
| PR | 01-DEC-2000 | 2000US-0250391 |
| PR | 05-DEC-2000 | 2000US-0251030 |
| PR | 05-DEC-2000 | 2000US-0251988 |
| PR | 06-DEC-2000 | 2000US-0256719 |
| PR | 06-DEC-2000 | 2000US-0251479 |
| PR | 08-DEC-2000 | 2000US-0251855 |
| PR | 08-DEC-2000 | 2000US-0251863 |
| PR | 08-DEC-2000 | 2000US-0251869 |
| PR | 08-DEC-2000 | 2000US-0251989 |
| PR | 11-DEC-2000 | 2000US-0251990 |
| PR | 05-JAN-2001 | 2000US-0259678 |

Rosen CA, Barash SC, Ruben SM,

WPI; 2001-465566/50.
M-PDND; 19941000

N-PSUB; AAS41090

prevent polypeptides and polynucleotides useful for diagnosing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases -

Claim 11; SEQ ID No 1216; 1180pp; English

The present invention relates to the isolation of novel human enzyme polypeptides, and the cDNA (AA540785-AA54184) and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS), autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophili), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. AAU22915-AAU23814 represent the novel human enzyme polypeptides of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

SQ Sequence 129 AA

| | | | | |
|--------------------------|--------|--------------------|-----------|-------------|
| Query Match | 34.9% | Score 469; | DB 22; | Length 129; |
| Best Local Similarity | 95.88; | Pred. No. 1.2e-40; | | |
| Matches 92; Conservative | 1; | Mismatches 3; | Indels 0; | Gaps 0 |

SO Sequence 146 AA;
Query Match 41.9%; Score 563; DB 22; Length 146;
Best Local Similarity 96.5%; Pred. No. 2.3e-50;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 HPETLVKVKDAEDQLGARVGYIELDLNSGELLEFRSEPRPMSTFKYLLCGAVLSRID 60
|||||
DB 28 HPETLVKVKDAEDQLGARVGYIELDLNSGELLEFRSEPRPMSTFKYLLCGAVLSRID 87
|||||

OY 61 AGGEOGLGRRIHYSONDLVEYSPVTEKHLTDGMTVRELCSAATMSNTANLLLT 115
|||||
DB 88 AGGEOGLGRRIHYSONDLVEYSPVTEKHLTDGMTVRELCSAATMSNTANLLLT 142
|||||

RESULT 4
AAW20440
ID AAW20440 standard; protein; 127 AA.
XX
AC AAW20440;
XX
DT 14-JUL-1997 (first entry)
XX
DE H. pylori cytoplasmic protein, 34574062.aa.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome;
KW replication; transcription; recombination; repair.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 121 /note= "encoded by codon YGC"
FT Misc-difference 122 /note= "encoded by codon WGC"
FT
XX
PN WO9640893-A1.
XX
PD 19-DEC-1996.
XX
PE 06-JUN-1996; 96MO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
XX (ASTR) ASTRA AB.
XX
PI Berglindh OT, Smith D, Mellgeerd BL;
XX
DR WPI; 1997-052306/05.
XX
DR N-PSDB; NAT67613.
XX
PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
XX
PS Claim 61; Pages 615-616; 1481pp; English.
XX
CC The present sequence is a Helicobacter pylori cytoplasmic protein
CC involved in genomic replication, transcription, recombination and repair.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds, useful
CC as potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
SQ Sequence 127 AA;
Query Match 36.0%; Score 484; DB 18; Length 127;
Best Local Similarity 97.9%; Pred. No. 3.2e-42;
Matches 95; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 HPETLVKVKDAEDQLGARVGYIELDLNSGELLEFRSEPRPMSTFKYLLCGAVLSRID 60
|||||
DB 24 HPETLVKVKDAEDQLGARVGYIELDLNSGELLEFRSEPRPMSTFKYLLCGAVLSRID 83
|||||

OY 61 AGGEOGLGRRIHYSONDLVEYSPVTEKHLTDGMTVREL 97
|||||
DB 84 AGGEOGLGRRIHYSONDLVEYSPVTEKHLTDGMTVREL 120
|||||

RESULT 5
AAU23220
ID AAU23220 standard; protein; 129 AA.
XX
AC AAU23220;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #306.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-02094515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.

PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226661.
 PR 22-AUG-2000; 2000US-0226868.
 PR 23-AUG-2000; 2000US-0227182.
 PR 30-AUG-2000; 2000US-0227009.
 PR 01-SEP-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234597.
 PR 25-SEP-2000; 2000US-0234598.
 PR 26-SEP-2000; 2000US-0234998.
 PR 27-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 29-SEP-2000; 2000US-0235835.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249247.
 PR 17-NOV-2000; 2000US-0249249.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR MPI: 2001-483426/52.
 XX N-PSDB; AAK63651.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Claim 11; SEQ ID NO 18463; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 CC
 XX

| | | |
|----|---|-----------------|
| PR | 20-OCT-2000; | 2000US-0241787. |
| PR | 20-OCT-2000; | 2000US-0241808. |
| PR | 20-OCT-2000; | 2000US-0241809. |
| PR | 20-OCT-2000; | 2000US-0241826. |
| PR | 01-NOV-2000; | 2000US-0244617. |
| PR | 08-NOV-2000; | 2000US-0246474. |
| PR | 08-NOV-2000; | 2000US-0246475. |
| PR | 08-NOV-2000; | 2000US-0246476. |
| PR | 08-NOV-2000; | 2000US-0246477. |
| PR | 08-NOV-2000; | 2000US-0246478. |
| PR | 08-NOV-2000; | 2000US-0246523. |
| PR | 08-NOV-2000; | 2000US-0246524. |
| PR | 08-NOV-2000; | 2000US-0246525. |
| PR | 08-NOV-2000; | 2000US-0246526. |
| PR | 08-NOV-2000; | 2000US-0246527. |
| PR | 08-NOV-2000; | 2000US-0246528. |
| PR | 08-NOV-2000; | 2000US-0246532. |
| PR | 08-NOV-2000; | 2000US-0246609. |
| PR | 08-NOV-2000; | 2000US-0246610. |
| PR | 08-NOV-2000; | 2000US-0246611. |
| PR | 08-NOV-2000; | 2000US-0246613. |
| PR | 17-NOV-2000; | 2000US-0249207. |
| PR | 17-NOV-2000; | 2000US-0249208. |
| PR | 17-NOV-2000; | 2000US-0249209. |
| PR | 17-NOV-2000; | 2000US-0249210. |
| PR | 17-NOV-2000; | 2000US-0249211. |
| PR | 17-NOV-2000; | 2000US-0249212. |
| PR | 17-NOV-2000; | 2000US-0249213. |
| PR | 17-NOV-2000; | 2000US-0249214. |
| PR | 17-NOV-2000; | 2000US-0249215. |
| PR | 17-NOV-2000; | 2000US-0249216. |
| PR | 17-NOV-2000; | 2000US-0249217. |
| PR | 17-NOV-2000; | 2000US-0249218. |
| PR | 17-NOV-2000; | 2000US-0249244. |
| PR | 17-NOV-2000; | 2000US-0249245. |
| PR | 17-NOV-2000; | 2000US-0249264. |
| PR | 17-NOV-2000; | 2000US-0249265. |
| PR | 17-NOV-2000; | 2000US-0249297. |
| PR | 17-NOV-2000; | 2000US-0249299. |
| PR | 17-NOV-2000; | 2000US-0249300. |
| PR | 01-DEC-2000; | 2000US-0250160. |
| PR | 01-DEC-2000; | 2000US-0250391. |
| PR | 05-DEC-2000; | 2000US-0251030. |
| PR | 05-DEC-2000; | 2000US-0251988. |
| PR | 05-DEC-2000; | 2000US-0256719. |
| PR | 06-DEC-2000; | 2000US-0251479. |
| PR | 08-DEC-2000; | 2000US-0251856. |
| PR | 08-DEC-2000; | 2000US-0251868. |
| PR | 08-DEC-2000; | 2000US-0251869. |
| PR | 08-DEC-2000; | 2000US-0251989. |
| PR | 08-DEC-2000; | 2000US-0251990. |
| PR | 11-DEC-2000; | 2000US-0254097. |
| PR | 05-JAN-2001; | 2001US-0259678. |
| PA | (HUMA-) HUMAN GENOME SCI INC. | |
| XX | | |
| PI | Rosen CA, Barash SC, Ruben SM; | |
| XX | | |
| DR | WPI; 2001-465566/50. | |
| DR | N-PSDB; AAS41091. | |
| XX | | |
| PT | Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases | |
| PT | - | |
| PS | Claim 11; SEQ ID NO 1217; 1180bp; English. | |
| XX | | |
| CC | The present invention relates to the isolation of novel human enzyme polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the | |
| CC | | |

| | |
|---------------------------|---|
| CC | diagnosis treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. |
| CC | AAM22915-AAU23814 represent the novel human enzyme polypeptides of the invention. |
| CC | Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. |
| CC | |
| XX | |
| XX | |
| SO | Sequence 146 AA: |
| | |
| Query Match | 41.9% Score 563; DB 22; Length 146; |
| Best local Similarity | 96.5% Pred. No. 2,3e-50; |
| Matches 111; Conservative | 1; Mismatches 3; Indels 0; Gaps 0; |
| OY | 1 HPEPTLVKVKAEEDLGARVGTIELDLNGCEILLESFRSEERFPFMASTFKVLICGAVLSRID 60 |
| DB | 28 HPETLVKVKAEEDLGARVGTYIELDLNSGKIIESEPFERFPMASTFKVLICGAVLSRID 87 |
| OY | 61 AGGEOLGRRIHYSONDLYEVSPTVEKHLLTDGMTVRELCSAATMSDNTAANILLT 115 |
| DB | 88 AGEOLGRRIHYSONDLYEVSPTVEKHLLTDGMTVRELCSAATMSDNTAANILLT 142 |
| RESULT 3 | |
| ID | AAM90870 |
| AC | AAM90870 standard; Protein: 146 AA. |
| XX | |
| XX | AAM90870; |
| DT | 07-NOV-2001 (first entry) |
| DE | |
| XX | Human immune/haematopoietic antigen SEQ ID NO:18463. |
| KW | Cytostatic; gene therapy; vaccine; metastasis. |
| OS | Homo sapiens. |
| PN | WO200157182-A2. |
| PD | |
| XX | 09-AUG-2001. |
| PF | |
| XX | 17-JAN-2001; 2001MO-USO1354. |
| PR | 31-JAN-2000; 2000US-0179065. |
| PR | 04-FEB-2000; 2000US-0180628. |
| PR | 24-FEB-2000; 2000US-0184664. |
| PR | 02-MAR-2000; 2000US-0186350. |
| PR | 16-MAR-2000; 2000US-0189874. |
| PR | 17-MAR-2000; 2000US-0190076. |
| PR | 18-APR-2000; 2000US-0198123. |
| PR | 19-MAY-2000; 2000US-0205515. |
| PR | 07-JUN-2000; 2000US-0209467. |
| PR | 28-JUN-2000; 2000US-0214886. |
| PR | 30-JUL-2000; 2000US-0215135. |
| PR | 07-JUL-2000; 2000US-0216647. |
| PR | 11-JUL-2000; 2000US-0216880. |
| PR | 11-JUL-2000; 2000US-0217487. |
| PR | 11-JUL-2000; 2000US-0217496. |
| PR | 14-JUL-2000; 2000US-0218290. |
| PR | 26-JUL-2000; 2000US-0220963. |
| PR | 26-JUL-2000; 2000US-0220964. |
| PR | 14-AUG-2000; 2000US-0224518. |
| PR | 14-AUG-2000; 2000US-0224519. |
| PR | 14-AUG-2000; 2000US-0225213. |
| PR | 14-AUG-2000; 2000US-0225214. |

PA (EART) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Komoto S, Nishimura A;
PI Matsushiro S;
XX
XX WPI, 1986-015031/03.
DR N-PSDB; AAN60531.
XX
PT New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PT and transformed cells contg. it.
XX
PS Disclosure: Page 56-59; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collects in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC Beta-urogastrone is the hormone of the salivary glands which suppresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obt. only
CC in small amts. from human urine.
CC See also AAN60628, and 30-32.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 196 AA;

Query Match 43.0%; Score 579; DB 7; Length 196;
Best Local Similarity 97.4%; Pred. No. 7.7e-52;
Matches 114; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 HPELVKVKADADQAGARVGYIELDNSGELLESFSESRERPPMSTFKYLGCAYLSRID 60
DB 24 HPELVKVKADADQAGARVGYIELDNSGELLESFSESRERPPMSTFKYLGCAYLSRID 83

OY 61 AGOEQLGRIRIHSQNDLVESPYTEKHLTDGMVRELCSAATMSDNTAANILFTTI 117
DB 84 AGOEQLGRIRIHSQNDLVESPYTEKHLTDGMVRELCSAATMSDNTAANILFTTI 140

RESULT 2
AAU23221
ID AAU23221 standard; Protein; 146 AA.
XX
XX AAU23221;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #307.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticogulant.
XX
OS Homo sapiens.
XX
XX WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 07-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239837.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E_37S

Perfect score: 1345

Sequence: 1 HPETLVKVKNAEDQGLGARVG.....TMDERRRQIAETGASLIKHW 263

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------------|--------------------|
| 1 | 579 | 43.0 | 196 | 7 AAP60627 | Beta-urogastrone - |
| 2 | 563 | 41.9 | 146 | 22 AAU23221 | Novel human enzyme |
| 3 | 563 | 41.9 | 146 | 22 AAM90870 | Human immune/haema |
| 4 | 484 | 36.0 | 127 | 18 AAW20440 | H. pylori cytoplas |
| 5 | 469 | 34.9 | 129 | 22 AAU23220 | Novel human enzyme |
| 6 | 469 | 34.9 | 129 | 22 AAM90871 | Human immune/haema |
| 7 | 463 | 34.4 | 182 | 22 ABG27917 | Novel human diagno |
| 8 | 463 | 34.4 | 94 | 21 AAB59052 | Breast and ovarian |
| 9 | 396 | 29.4 | 159 | 7 AAP60628 | Beta-urogastrone - |

| | | | | | |
|----|------|------|-----|-------------|---------------------|
| 10 | 372 | 27.7 | 101 | 22 ABG27935 | Novel human diagno |
| 11 | 325 | 24.2 | 88 | 22 ABG27919 | Novel human diagno |
| 12 | 191 | 14.2 | 119 | 7 AAP60626 | Beta-urogastrone - |
| 13 | 141 | 10.5 | 51 | 21 AAY92783 | Vtgs-beta-lactama |
| 14 | 81 | 6.0 | 200 | 18 AAM55524 | H. pylori ORF 29ep |
| 15 | 79.5 | 5.9 | 158 | 22 AAU45138 | Protonibacterium |
| 16 | 78 | 5.8 | 202 | 24 AAE32763 | Human zcyto24 prot |
| 17 | 78 | 5.8 | 202 | 24 AAO16283 | Mouse IXX129840-2 |
| 18 | 71 | 5.3 | 159 | 19 AAM68170 | N-terminally tagge |
| 19 | 70 | 5.2 | 15 | 10 AAP98503 | Sequence encoded b |
| 20 | 70 | 5.2 | 202 | 22 AAG82686 | S. epidermidis ope |
| 21 | 69.5 | 5.1 | 170 | 22 AAU56330 | Protonibacterium |
| 22 | 69 | 5.1 | 134 | 21 AAG27220 | Zea mays protein f |
| 23 | 69 | 5.1 | 178 | 22 AAG91926 | C glutamincum prote |
| 24 | 69 | 5.1 | 180 | 21 AAG20206 | Arabidopsis thalia |
| 25 | 69 | 5.1 | 194 | 22 AAG81764 | S. epidermidis ope |
| 26 | 67.5 | 5.0 | 177 | 21 AAB41582 | Human ORFX ORF1345 |
| 27 | 67.5 | 5.0 | 184 | 21 AAB25446 | Pinus radiata cell |
| 28 | 67.5 | 5.0 | 187 | 23 AAU97104 | Human MK61 protein |
| 29 | 67 | 5.0 | 101 | 22 AAU66278 | Protonibacterium |
| 30 | 67 | 5.0 | 168 | 22 AAB68717 | Drosophila melanog |
| 31 | 67 | 5.0 | 202 | 24 AAE32764 | Human zcyto25 prot |
| 32 | 66.5 | 4.9 | 196 | 21 AAY73344 | HTRM clone 0258181 |
| 33 | 66 | 4.9 | 170 | 22 AAU52472 | Protonibacterium |
| 34 | 65.5 | 4.9 | 50 | 17 AAW03591 | Human alpha 2 C4 a |
| 35 | 65.5 | 4.9 | 102 | 23 AAB34479 | Human dehydrogenas |
| 36 | 65.5 | 4.9 | 136 | 21 AAY44229 | Partial corn extra |
| 37 | 65.5 | 4.9 | 152 | 22 ABG19138 | Novel human diagno |
| 38 | 65.5 | 4.9 | 152 | 22 ABG27771 | Novel human diagno |
| 39 | 65.5 | 4.9 | 170 | 19 AAM38722 | S. pneumoniae SGHR |
| 40 | 65.5 | 4.9 | 191 | 22 AAU87418 | Novel central nerv |
| 41 | 65.5 | 4.9 | 191 | 22 AAU19740 | Human novel extrinc |
| 42 | 65.5 | 4.9 | 191 | 23 AAB47960 | Human polypeptide |
| 43 | 65 | 4.8 | 113 | 22 AAB49942 | Protein disulfide |
| 44 | 65 | 4.8 | 174 | 16 AAB6508 | Aspergillus sp. tr |
| 45 | 65 | 4.8 | 179 | 22 AAU87627 | Novel central nerv |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AAU60627 | |
| ID | AAU60627 standard; Protein: 196 AA. |
| XX | |
| AC | AAU60627; |
| XX | |
| DT | 25-MAR-2003 (updated) |
| DT | 17-JUN-1991 (first entry) |
| XX | |
| DE | Beta-urogastrone - beta-lactamase fusion protein from pUG2101. |
| XX | |
| KW | Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds; |
| KW | fusion protein; beta-lactamase. |
| XX | |
| OS | Synthetic. |
| XX | |
| FH | Key |
| FT | Protein |
| FT | Location/Qualifiers |
| FT | 1..120 |
| FT | /label= beta-lactamase |
| FT | Misc-difference 121..123 |
| FT | /label= adaptor |
| FT | Protein |
| FT | 124..196 |
| FT | /label= beta-urogastrone |
| XX | |
| PN | DE3523634-A. |
| XX | |
| PD | 09-JAN-1986. |
| XX | |
| PF | 02-JUL-1985; |
| XX | 85DE-3523634. |
| PR | 02-JUL-1984; |
| XX | 84JP-0137691. |
| XX | |

Novel human diagno
Novel human diagno
Beta-urogastrone -
Vtgs-beta-lactama
H. pylori ORF 29ep
Protonibacterium
Human zcyto24 prot
Mouse IXX129840-2
N-terminally tagge
Sequence encoded b
S. epidermidis ope
Protonibacterium
Zea mays protein f
C glutamincum prote
Arabidopsis thalia
S. epidermidis ope
Human ORFX ORF1345
Pinus radiata cell
Human MK61 protein
Protonibacterium
Drosophila melanog
Human zcyto25 prot
HTRM clone 0258181
Protonibacterium
Human alpha 2 C4 a
Human dehydrogenas
Partial corn extra
Novel human diagno
Novel human diagno
S. pneumoniae SGHR
Novel central nerv
Human novel extrinc
Human polypeptide
Protein disulfide
Aspergillus sp. tr
Novel central nerv

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_157T
Perfect score: 1348
Sequence: 1 HPETLVKKAEDQIGARVG.....TMDERNQIAICASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacterioid:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|-------------|
| 1 | 524 | 38.9 | 105 | 2 | 052026 |
| 2 | 524 | 38.9 | 105 | 2 | 052330 |
| 3 | 499 | 37.0 | 145 | 2 | 08RTD8 |
| 4 | 488 | 36.2 | 102 | 2 | 052639 |
| 5 | 464 | 34.4 | 138 | 2 | 099QF3 |
| 6 | 464 | 34.4 | 138 | 2 | 09AMA1 |
| 7 | 464 | 34.4 | 139 | 2 | 09AMA2 |
| 8 | 462 | 34.3 | 138 | 2 | 09AMA0 |
| 9 | 462 | 34.3 | 138 | 2 | 09AMA9 |
| 10 | 462 | 34.3 | 139 | 2 | 09AMA9 |
| 11 | 459 | 34.1 | 139 | 2 | 09AMA3 |
| 12 | 338 | 25.1 | 67 | 2 | 053553 |
| 13 | 326 | 24.2 | 95 | 2 | 08GDE5 |
| 14 | 310 | 23.0 | 62 | 2 | 08GDE5 |
| 15 | 301 | 22.3 | 128 | 2 | 08V000 |
| 16 | 251 | 18.6 | 180 | 2 | 08KVT2 |

| | | | | | | |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 250 | 18.5 | 52 | 2 | 09R412 | 09r412 shigella fl |
| 18 | 241 | 17.9 | 48 | 2 | 09RLH0 | 09rlh0 proteus mtr |
| 19 | 197 | 14.6 | 38 | 2 | P97145 | P97145 escherichia |
| 20 | 168.5 | 12.5 | 134 | 2 | 08VUL3 | 08vul3 staphylococ |
| 21 | 123.5 | 9.2 | 109 | 2 | 053698 | 053698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | P97146 | P97146 escherichia |
| 23 | 87 | 6.5 | 100 | 2 | 093S05 | 093s05 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | 09XBJ2 | 09xbj2 bacillus ce |
| 25 | 80 | 5.9 | 198 | 2 | 09ACM8 | 09acm8 streptococ |
| 26 | 79.5 | 5.9 | 202 | 16 | 09S520 | 09s520 rhizobium l |
| 27 | 77 | 5.7 | 153 | 5 | 08SS20 | 08ss20 yersinia pe |
| 28 | 75 | 5.6 | 181 | 16 | 08D108 | 08d108 mycobacteri |
| 29 | 74.5 | 5.5 | 113 | 2 | 049870 | 049870 brucella me |
| 30 | 74.5 | 5.5 | 172 | 16 | 08RTW3 | 08rtw3 yersinia en |
| 31 | 74.5 | 5.5 | 172 | 16 | 08RTW3 | 08rtw3 yersinia en |
| 32 | 73 | 5.4 | 131 | 2 | 09X9H0 | 09x9h0 yersinia en |
| 33 | 73 | 5.4 | 175 | 17 | 08ZTH8 | 08zth8 pyrobaculum |
| 34 | 72.5 | 5.4 | 145 | 2 | 005984 | 005984 staphylococ |
| 35 | 72.5 | 5.4 | 204 | 11 | 091YW4 | 091yw4 mus musculu |
| 36 | 72 | 5.3 | 205 | 16 | 09BJ18 | 09bj18 rhizobium l |
| 37 | 71.5 | 5.3 | 152 | 16 | 09RI91 | 09ri91 streptomyce |
| 38 | 71.5 | 5.3 | 190 | 5 | 076227 | 076227 trypanosoma |
| 39 | 71.5 | 5.3 | 195 | 16 | 09RIA0 | 09ria0 streptomyc |
| 40 | 71.5 | 5.3 | 196 | 2 | 09AH34 | 09ah34 pseudomonas |
| 41 | 71.5 | 5.3 | 200 | 16 | 097JEB | 097jeb clostridium |
| 42 | 71 | 5.3 | 150 | 10 | 094IG4 | 094ig4 nicotiana t |
| 43 | 71 | 5.3 | 153 | 16 | 09PKT4 | 09pkt4 chlamydia t |
| 44 | 71 | 5.3 | 177 | 16 | 09KEM7 | 09kem7 bacillus ha |
| 45 | 71 | 5.3 | 192 | 16 | 09HW07 | 09hw07 pseudomonas |

ALIGNMENTS

RESULT 1

| | | | | |
|--|---|--------------|------|---------|
| ID | 052026 | PRELIMINARY: | PRT: | 105 AA. |
| AC | 052026: | | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Created) | | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Last sequence update) | | | |
| DT | 01-MAR-2002 (TREMBLrel. 20, Last annotation update) | | | |
| DE | Beta-lactamase (Fragment). | | | |
| GN | BLA. | | | |
| OS | Pseudomonas aeruginosa. | | | |
| OC | Plasmid PR01614. | | | |
| OC | Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales; | | | |
| OC | Pseudomonadaceae; Pseudomonas. | | | |
| OX | NCBI_TaxID=287; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=95011664; PubMed=7926843; | | | |
| RA | West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.; | | | |
| RT | *Construction of Improved Escherichia-Pseudomonas Shuttle Vectors | | | |
| RT | derived from pUC18/19 and sequence of the region required for their | | | |
| RT | replication in Pseudomonas aeruginosa.; | | | |
| RL | Gene 148:81-86(1994). | | | |
| DR | EMBL: L30112; AAA6058.1; - | | | |
| DR | HSSP: P00810; 1XPB. | | | |
| DR | InterPro: IPR001466; Beta_Lactamase. | | | |
| DR | InterPro: IPR000871; Beta_Lactamase_A. | | | |
| DR | Pfam: PF00144; Beta_Lactamase; 1. | | | |
| DR | PRINTS: PR00118; BLACTAMASRA. | | | |
| FT | Plasmid. | | | |
| FT | NON_TER | | | |
| SQ | SEQUENCE | 1 | 1 | |
| | 105 AA; 11229 MW; D2889A4073330557 CRC64; | | | |
| Query Match | 38.9%; Score 524; DB 2; Length 105; | | | |
| Best Local Similarity | 100.0%; Pred. No. 8.5e-37; | | | |
| Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | | | |
| QY | 160 AMATTIKRLTGLTLLASROOLIMWEADKRVAGPILRSALPAGWFLADKSGGERSRG 219 | | | |
| DB | 2 AMATTIKRLTGLTLLASROOLIMWEADKRVAGPILRSALPAGWFLADKSGGERSRG 61 | | | |

OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNQIAEIGASLIKHW 263
 DB 62 IIAALGPDGKPSRIIVYITGSOATMDERNQIAEIGASLIKHW 105

RESULT 2

OY 052330 PRELIMINARY; PRT; 105 AA.
 AC 052330;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Tnl bla protein (fragment).
 OS Escherichia coli.
 OC Plasmid RK2.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;

RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE-90264294; PubMed-2160936;
 RA Kornacki J.A., Burlage R.S., Figueroa D.H.;
 RT "The *kil*-kor regulation of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT korc.";
 RL J. Bacteriol. 172:3040-3050(1990).
 DR EMBL: M32794; AAA26408.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR InterPro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER 1 1
 SO SEQUENCE 105 AA; 11229 MW; D2889A407330557 CRC64;

Query Match 38.9%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 8.5e-37;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AAMATIRKLITGELTILASRQQLIDMWEADKVAAGPLRSALPAGFIADKSGAGERSRG 219
 DB 2 AAMATIRKLITGELTILASRQQLIDMWEADKVAAGPLRSALPAGFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNQIAEIGASLIKHW 263
 DB 62 IIAALGPDGKPSRIIVYITGSOATMDERNQIAEIGASLIKHW 105

RESULT 3

OY 08RTD8 PRELIMINARY; PRT; 145 AA.
 AC 08RTD8;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE SHV-5 enzyme (fragment).
 OS Klebsiella pneumoniae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;

RN (1)
 RP SEQUENCE FROM N.A.
 RA Palisubramaniam S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT cefazidime-resistant *Klebsiella pneumoniae*,";
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF467105; AAL75506.1; -.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER 1 1
 SO SEQUENCE 145 AA; 145

SO SEQUENCE 145 AA; 15574 MW; F8863406194B4C82 CRC64;
 Query Match 37.0%; Score 499; DB 2; Length 145;
 Best Local Similarity 68.8%; Pred. No. 1.8e-34;
 Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

OY 100 AATMSDNTAANLLTTTIGPKELTAFLHNMGDHTRDRPELNEALPDERDTPPV 159
 DB 7 AAXYSDNSAANLLATVGGPAGLTAFLRQIGDVTIRDRMETLNEALPGDARDTPPA 66
 OY 160 AAMATIRKLITGELTILASRQQLIDMWEADKVAAGPLRSALPAGFIADKSGAGERSRG 219
 DB 67 SMAATIRKLITGELTILASRQQLIDMWEADKVAAGPLRSALPAGFIADKSGAGERSRG 126
 OY 220 IIAALGPDGKPSRIIVY 237
 DB 127 IVALGPNKKAERIVY 144

RESULT 4

OY 052639 PRELIMINARY; PRT; 102 AA.
 AC 052639;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Beta-lactamase (fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 CC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;

RN (1)
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-Tn1.
 RX MEDLINE-94336757; PubMed-8058819;
 RA Jansons I., Touchle G., Sharp R., Almqvist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT "Deletion and transposon mutagenesis and sequence analysis of the *pdr*
 RT *PRO1600* *Orf1* region found in the broad-host-range plasmids of the *pdr*
 RT series.";
 RL Plasmid 31:265-274(1994).
 DR EMBL: L22691; AAA98312.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_Lactamase.
 DR InterPro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER 1 1
 SO SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.2%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 9e-34;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AAMATIRKLITGELTILASRQQLIDMWEADKVAAGPLRSALPAGFIADKSGAGERSRG 219
 DB 2 AAMATIRKLITGELTILASRQQLIDMWEADKVAAGPLRSALPAGFIADKSGAGERSRG 61
 OY 220 IIAALGPDGKPSRIIVYITGSOATMDERNQIAEIGA 257
 DB 62 IIAALGPDGKPSRIIVYITGSOATMDERNQIAEIGA 99

RESULT 5

OY 0990F3 PRELIMINARY; PRT; 138 AA.
 AC 0990F3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

RC STRAIN-E/98 9-1;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327052; AK07467.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15105 MW; 561D092E5442A847 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 2.2e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFLHNGDHYTRLDWEPELNEAIPNDERDTTPVAMATTLKRLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNTRLDRWETELNEALPGDARDTTPASMAATLRLKLT 69
QY 171 GELLTLASROOLIDMEADKVAAGPLRSALPAGWFIADKSGAGSGSGIITAAIGPDGR 230
DB 70 SQRLSARSOQLQWVDDRVAGPLIRSVLPAGWFIADKSGAGSGIITAAIGPDGR 129
QY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 9

Q9AM99 PRELIMINARY; PRT; 138 AA.

AC Q9AM99: 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/98 4-1;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327053; AK07468.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 2.2e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFLHNGDHYTRLDWEPELNEAIPNDERDTTPVAMATTLKRLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNTRLDRWETELNEALPGDARDTTPASMAATLRLKLT 69
QY 171 GELLTLASROOLIDMEADKVAAGPLRSALPAGWFIADKSGAGSGSGIITAAIGPDGR 230
DB 70 SQRLSARSOQLQWVDDRVAGPLIRSVLPAGWFIADKSGAGSGIITAAIGPDGR 129

QY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 10

Q9AM98 PRELIMINARY; PRT; 139 AA.

AC Q9AM98: 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/98 3-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327054; AK07469.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam: PF00144; beta-lactamase_A.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15234 MW; 0361A792E5442A8 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 2.2e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKELTAFLHNGDHYTRLDWEPELNEAIPNDERDTTPVAMATTLKRLT 170
DB 10 NLLATVGGPAGLTAFLRQIGDNTRLDRWETELNEALPGDARDTTPASMAATLRLKLT 69
QY 171 GELLTLASROOLIDMEADKVAAGPLRSALPAGWFIADKSGAGSGSGIITAAIGPDGR 230
DB 70 SQRLSARSOQLQWVDDRVAGPLIRSVLPAGWFIADKSGAGSGIITAAIGPDGR 129
QY 231 SRIVIVY 237
DB 130 ERIVIVY 136

RESULT 11

Q9AM93 PRELIMINARY; PRT; 139 AA.

AC Q9AM93: 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-935;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
RL Isolates from UMMC, Malaysia.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327047; AK07462.1; -.
DR HSSP: P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.

```
DR InterPro: IPR000871; Beta-lactamase-A.
DR Pfam: PF00144; beta-lactamase: 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 1 1
FT NON_TER 139 139
SQ SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Query Match
Best Local Similarity 67.7%; Score 459; DB 2; Length 139;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

QY 111 NLLTTTGGPKETLAFVLMNGDHTRLDRWPEPELNEALPNDERDTTPVAMATTLRLTLT 170
DB 10 HLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTTASMAATLRLTLT 69
QY 171 GELLVLASRQQLIDWMEADRVAPLRSALPAGFTADKSGAGRGSGITIAALGPDGXP 230
DB 70 SORLSARSRQLLQWVDDRVAGPLRLSVLPAGWFIADKTSKRGAGIYALLGPNNKA 129
QY 231 SRIVVYIY 237
DB 130 ERIVVYIY 136

RESULT 12
Q53553 PRELIMINARY; PRT; 67 AA.
AC 053553.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_Taxid=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE-96081517; PubMed-8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL; S81098; AAB35839.2; -.
DR HSSP; P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase: 1.
DR PRINTS; PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 67 67
SQ SEQUENCE 67 AA; 7287 MW; 85F8C85B0363F0CB CRC64;

Query Match
Best Local Similarity 25.1%; Score 338; DB 2; Length 67;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 55 VLSHIDAGQEQGLGRIRHSQNDLVEYSPYTERKHLTDGMTVRELCSAATNSDNTAANLL 114
DB 1 VLSHVDAQEQQLGRIRHSQNDLVEYSPYTERKHLTDGMTVRELCSAATNSDNTAANLL 60
QY 115 TTITGPK 121
DB 61 TTITGPK 67

RESULT 13
Q8GDES PRELIMINARY; PRT; 95 AA.
AC 08GDES;
DB 130 AG 193
```

```
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_Taxid=573;
RN [1]
RP SEQUENCE FROM N.A.
RA Bello H.M., Dominguez M.P., Dashti A.A., Gonzalez-Rocha G.E.,
RA Amey S.G.B.;
RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
RT Chile."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY150585; AAN77730.1; -.
FT NON_TER 1 1
FT NON_TER 95 95
SQ SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
Best Local Similarity 24.2%; Score 326; DB 2; Length 95;
Matches 62; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 147 AIPNDERDTTPVAMATTLRLTLTGELTLASRQQLIDWMEADRVAGPLRSALPAGFTI 206
DB 1 ALPGARDTTTPASMAATLRLTLTSORLSARSRQLLQWVDDRVAGPLRLSVLPAGFTI 60
QY 207 ADKSGAGRGSGRGTIAALGPDGKPSRIVYIYTGGS 241
DB 61 ADKTGASERGAGIYALLGPNNKAERIVYIYLRDS 95

RESULT 14
Q9JN58 PRELIMINARY; PRT; 62 AA.
ID Q9JN58.
AC Q9JN58.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE TEM-type beta-lactamase (Fragment).
OS Shigella flexneri.
OC Plasmid pMAM-1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_Taxid=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-UCSM 129;
RX MEDLINE-96081517; PubMed-8538491;
RA Echeverria V., Olate J., Cid H.;
RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
RT flexneri strain."
RL Microbios 83:107-117(1995).
DR EMBL; S81099; AAB35840.1; -.
DR HSSP; P00810; 1XPB.
DR InterPro: IPR000871; Beta_lactamase_A.
DR PRINTS; PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER 1 1
FT NON_TER 62 62
SQ SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
Best Local Similarity 23.0%; Score 310; DB 2; Length 62;
Matches 60; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 132 DHTRLDRWPEPELNEALPNDERDTTPVAMATTLRLTLTGELTLASRQQLIDWMEADRV 191
DB 1 DHTRLDRWPEPELNEALPNDERDTTPVAMATTLRLTLTGELTLASRQQLIDWMEADRV 60
QY 192 AG 193
```

Db 61 AC 62

RESULT 15

08VQ00

| | | | | |
|----|--------|--------------|------|---------|
| ID | QBVQ00 | PRELIMINARY; | PRT; | 128 AA. |
|----|--------|--------------|------|---------|

AC QBVQ00;

DT 01-MAR-2002 (TREMBLREL, 20, Created)

DT 01-MAR-2002 (Tremblay, 20, Last sequence update)

DT 01-MAR-2003 (Tremblay, 23, last annotation update)

DE CIX-M type beta-lactamase (Fragment)

OS Escherichia coli.
OC Bacteria: Protob

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia

OC Enterobacteriaceae; Escherichia.
CY NCBI TaxID=562.

OX NCBI_TaxID=562;
PN [1]

RN (11)
RP SEQUENCE FROM N. A.

| RP | SEQUENCE FROM N.A. |
|----|--------------------|
| RC | STRAIN-clinical 19 |

RA Stuerenburg E.: Feucht H.: I

RT "A new CTX-M type beta-lactamase.";
KA Stuebenburg E.; Feuchil H.; Louis A.;
KA

Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF466148; MML73983.1; -

DR InterPro; IPR001466; Beta_lactamase.

DR InterPro; IPR000871; Beta_lactamase_A.

DR Pfam; PF00144; beta-lactamase; 1.

DR PRINTS; PR00118; BLACTAMASEA.

| | | | |
|----|---------|---|---|
| | NON_TER | 1 | 1 |
| ET | | | |

| | | | |
|----|-------------|-----|-----|
| FT | NON_TER | 128 | 128 |
| CO | COEFFICIENT | 128 | 128 |

50 SEQUENCE 128 AA; 13878 MW; F2EB815BB4FF318A CRC64;

Quarry Match 22 38. Score 301. DA 2. Length 128.

| | | | |
|-----------------------|-------|------------|-----------|
| Query Match | 22.3% | Score 301; | DB 2; |
| Best Local Similarity | 48.4% | Pred. NO. | 8 Jan-18: |

Best Local Similarity 48.4%; Pred. NO. 8.1e-18;
Matches 59: Conservative 20: Mismatches 43: Indels 0: Gaps 0:

QY 66 LGRRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAITMSDNTAANLLTTIGPKELTA 125

[illegible]

126 FLHNMGDHVTRLDRWEPELNEAIPNDERDTTPVAMATTLRKLLTGELLTLASRQQLDW 185

— — — — —

186 ME 187
ay

— 2 —

Search completed: September 10, 2003, 12:29:27
Job time : 25.4286 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E
Perfect score: 1348
Sequence: 1 HPETLVKVKDAEDQLGARVY.....TMDERNRIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 587 | 43.5 | 196 | 7 | AA60627 |
| 2 | 571 | 42.4 | 146 | 22 | AAU23221 |
| 3 | 571 | 42.4 | 146 | 22 | AAU23221 |
| 4 | 492 | 36.5 | 127 | 18 | AAW20440 |
| 5 | 477 | 35.4 | 129 | 22 | AAU23220 |
| 6 | 477 | 35.4 | 129 | 22 | AAU23220 |
| 7 | 477 | 35.4 | 129 | 22 | AAU23220 |
| 8 | 463 | 34.3 | 94 | 21 | AAU23220 |
| 9 | 404 | 30.0 | 159 | 7 | AA60628 |

| | | | | | |
|----|------|------|-----|----|-----------|
| 10 | 372 | 27.6 | 101 | 22 | ABG27935 |
| 11 | 333 | 24.7 | 88 | 22 | ABG27919 |
| 12 | 199 | 14.8 | 119 | 7 | AA60626 |
| 13 | 141 | 10.5 | 51 | 21 | AAU232783 |
| 14 | 79.5 | 5.9 | 158 | 22 | AAU45138 |
| 15 | 79.5 | 5.9 | 200 | 18 | AAU55524 |
| 16 | 78 | 5.8 | 15 | 10 | AAU98503 |
| 17 | 78 | 5.8 | 202 | 24 | AAU32763 |
| 18 | 78 | 5.8 | 202 | 24 | AAU16283 |
| 19 | 73.5 | 5.5 | 170 | 19 | AAU38722 |
| 20 | 71 | 5.3 | 159 | 19 | AAU69170 |
| 21 | 70 | 5.2 | 202 | 22 | AAU62586 |
| 22 | 69.5 | 5.2 | 170 | 22 | AAU65330 |
| 23 | 69 | 5.1 | 134 | 21 | AAU27220 |
| 24 | 69 | 5.1 | 180 | 21 | AAU91926 |
| 25 | 69 | 5.1 | 180 | 21 | AAU91926 |
| 26 | 69 | 5.1 | 194 | 22 | AAU81764 |
| 27 | 67.5 | 5.0 | 184 | 21 | AAU25446 |
| 28 | 67.5 | 5.0 | 187 | 21 | AAU605495 |
| 29 | 67.5 | 5.0 | 187 | 21 | AAU605495 |
| 30 | 67.5 | 5.0 | 187 | 21 | AAU605495 |
| 31 | 67 | 5.0 | 101 | 22 | AAU66278 |
| 32 | 67 | 5.0 | 168 | 22 | AAU66278 |
| 33 | 67 | 5.0 | 202 | 24 | AAU32764 |
| 34 | 66 | 4.9 | 170 | 22 | AAU52472 |
| 35 | 66 | 4.9 | 196 | 18 | AAU14564 |
| 36 | 65.5 | 4.9 | 102 | 17 | AAU03591 |
| 37 | 65.5 | 4.9 | 102 | 17 | AAU03591 |
| 38 | 65.5 | 4.9 | 136 | 21 | AAU44429 |
| 39 | 65.5 | 4.9 | 152 | 22 | AAU19138 |
| 40 | 65.5 | 4.9 | 152 | 22 | AAU27771 |
| 41 | 65.5 | 4.9 | 177 | 21 | AAU41582 |
| 42 | 65.5 | 4.9 | 191 | 22 | AAU87418 |
| 43 | 65.5 | 4.9 | 191 | 22 | AAU19740 |
| 44 | 65.5 | 4.9 | 191 | 22 | AAU19740 |
| 45 | 65.5 | 4.9 | 196 | 21 | AAU73344 |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AA60627 | standard; Protein; 196 AA. |
| XX | |
| AC | AA60627: |
| XX | |
| DT | 25-MAR-2003 (updated) |
| DT | 17-JUN-1991 (first entry) |
| XX | |
| DE | Beta-urogastrone - beta-lactamase fusion protein from pUG2101. |
| XX | |
| KW | Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds; |
| KW | fusion protein; beta-lactamase. |
| XX | |
| OS | Synthetic. |
| XX | |
| FT | Key |
| FT | Protein |
| FT | Misc-difference |
| FT | Protein |
| FT | Protein |
| XX | |
| PN | DE3523634-A. |
| XX | |
| PD | 09-JAN-1986. |
| XX | |
| PF | 02-JUL-1985; 85DE-3523634. |
| XX | |
| PR | 02-JUL-1984; 84JP-0137691. |
| XX | |

Novel human diagno
Novel human diagno
Beta-urogastrone -
Vegs-beta-lactama
Proprionibacterium
H. pylori ORF 29ep
Sequence encoded b
Human zcyto24 prot
Mouse IKX129840-2
S. pneumoniae SGHR
N-terminally tagge
S. epidermidis ope
Proprionibacterium
Zea mays protein f
C glutamium prote
Arabidopsis thalia
S. epidermidis ope
Pinus radiata cell
Arabidopsis thalia
Human MR61 protein
Proprionibacterium
Drosophila melanog
Human zcyto25 prot
Streptococcus pneu
Human alpha 2 C4 a
Human denhydrogen
Partial corn extra
Novel human diagno
Novel human diagno
Human ORF ORF1346
Novel central nerv
Human novel extrac
Human polypeptide
HTRW clone 0258181

```

PA (EART ) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgai H, Hozinaka A, Hiramatsu H, Koumoto S, Nishimura A;
PI Matsushiro S;
XX
XX WPI: 1986-015031/03.
XX
XX N-PSDB; AAN60631.
XX
XX
XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
PT and transformed cells contg. It.
XX
XX
XX Disclosure; Page 56-59; 92pp; German.
XX
XX
XX The fusion protein is less easily degraded by proteases and so
CC protects beta-urogastrone and beta-lactamase collectis in the periplasm
CC of E.coli. It is therefore easy to collect and purify the product.
CC
CC Beta-urogastrone is the hormone of the salivary glands which supresses
CC stomach acid secretion and promotes cell growth, so is useful for
CC treating ulcers and wounds. Previously the product was obtd. only
CC in small amts. from human urine.
CC
CC See also AAN60628, and 30-32.
CC
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX
XX Sequence 196 AA;
SO
Query Match 43.5%; Score 587; DB 7; Length 196;
Best Local Similarity 98.3%; Pred. NO. 3.3e-53;
Matches 115; Conservative 2; Mismatches 0; Indels 0; Gaps 0
OY 1 HPEETLVKVKDAEDOLGARVGITELDLNNGEILSEFPERFERPMSTFEKVLICGAVLSRID 60
DB 24 HPEETLVKVKDAEDOLGARVGITELDLNNGSKILESFRPERFERPMSTFEKVLICGAVLSRVD 83
OY 61 AGGEQLGRRIRHYSONDLVEYSPYTEKHLTDGMTVRELCSAATMSDNTAANLLTTI 117
DB 84 AGGEQLGRRIRHYSONDLVEYSPYTEKHLTDGMTVRELCSAATMSDNTAANLLTTI 140
RESULT 2
AAU23221
ID AAU23221 standard; Protein: 146 AA.
XX
XX AAU23221;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #307.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
XX autoimmune disorder; neurological disorder; metabolic disorder;
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX blood-related disorder; infectious disorder; cystostatic; anti arthritic;
XX nephrotoxic; anticoagulant.
XX
XX Homo sapiens.
XX
XX WO200155301-A2.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001MO-US01239.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX

```

| | | |
|----|-------------|------------------|
| PR | 07-JUN-2000 | 200005-021513.55 |
| PR | 07-JUL-2000 | 200005-021664.7 |
| PR | 07-JUL-2000 | 200005-021688.0 |
| PR | 11-JUL-2000 | 200005-021744.87 |
| PR | 11-JUL-2000 | 200005-021749.6 |
| PR | 14-JUL-2000 | 200005-021829.0 |
| PR | 26-JUL-2000 | 200005-022066.3 |
| PR | 26-JUL-2000 | 200005-022096.4 |
| PR | 26-JUL-2000 | 200005-022451.8 |
| PR | 14-AUG-2000 | 200005-022451.9 |
| PR | 14-AUG-2000 | 200005-022451.9 |
| PR | 14-AUG-2000 | 200005-022523.13 |
| PR | 14-AUG-2000 | 200005-022523.14 |
| PR | 14-AUG-2000 | 200005-022523.65 |
| PR | 14-AUG-2000 | 200005-022523.67 |
| PR | 14-AUG-2000 | 200005-022526.68 |
| PR | 14-AUG-2000 | 200005-022526.79 |
| PR | 18-AUG-2000 | 200005-022632.79 |
| PR | 22-AUG-2000 | 200005-022668.68 |
| PR | 22-AUG-2000 | 200005-022710.92 |
| PR | 23-AUG-2000 | 200005-022710.92 |
| PR | 30-AUG-2000 | 200005-022892.4 |
| PR | 01-SEP-2000 | 200005-022829.87 |
| PR | 01-SEP-2000 | 200005-022833.43 |
| PR | 01-SEP-2000 | 200005-022833.44 |
| PR | 01-SEP-2000 | 200005-022833.45 |
| PR | 05-SEP-2000 | 200005-022850.9 |
| PR | 05-SEP-2000 | 200005-022851.3 |
| PR | 06-SEP-2000 | 200005-023043.7 |
| PR | 06-SEP-2000 | 200005-023043.78 |
| PR | 08-SEP-2000 | 200005-023134.22 |
| PR | 08-SEP-2000 | 200005-023134.23 |
| PR | 08-SEP-2000 | 200005-023124.4 |
| PR | 08-SEP-2000 | 200005-023141.3 |
| PR | 08-SEP-2000 | 200005-023141.4 |
| PR | 08-SEP-2000 | 200005-023308.10 |
| PR | 08-SEP-2000 | 200005-023308.14 |
| PR | 12-SEP-2000 | 200005-023308.1 |
| PR | 12-SEP-2000 | 200005-023339.7 |
| PR | 14-SEP-2000 | 200005-023339.9 |
| PR | 14-SEP-2000 | 200005-023339.9 |
| PR | 14-SEP-2000 | 200005-023440.1 |
| PR | 14-SEP-2000 | 200005-023440.1 |
| PR | 14-SEP-2000 | 200005-023306.3 |
| PR | 14-SEP-2000 | 200005-023306.4 |
| PR | 14-SEP-2000 | 200005-023306.5 |
| PR | 21-SEP-2000 | 200005-023422.74 |
| PR | 21-SEP-2000 | 200005-023422.74 |
| PR | 25-SEP-2000 | 200005-023499.7 |
| PR | 25-SEP-2000 | 200005-023499.8 |
| PR | 26-SEP-2000 | 200005-023548.4 |
| PR | 27-SEP-2000 | 200005-023583.4 |
| PR | 27-SEP-2000 | 200005-023583.6 |
| PR | 29-SEP-2000 | 200005-023632.7 |
| PR | 29-SEP-2000 | 200005-023632.7 |
| PR | 29-SEP-2000 | 200005-023636.8 |
| PR | 29-SEP-2000 | 200005-023636.9 |
| PR | 29-SEP-2000 | 200005-023637.0 |
| PR | 02-OCT-2000 | 200005-023660.2 |
| PR | 02-OCT-2000 | 200005-023703.8 |
| PR | 02-OCT-2000 | 200005-023703.8 |
| PR | 02-OCT-2000 | 200005-023703.9 |
| PR | 02-OCT-2000 | 200005-023704.0 |
| PR | 13-OCT-2000 | 200005-023933.5 |
| PR | 13-OCT-2000 | 200005-023933.7 |
| PR | 20-OCT-2000 | 200005-024096.0 |
| PR | 20-OCT-2000 | 200005-024122.1 |
| PR | 20-OCT-2000 | 200005-024178.5 |
| PR | 20-OCT-2000 | 200005-024178.6 |

20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246510.
PR 08-NOV-2000; 2000US-0246511.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PA Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465566/50.
DR N-PSDB; AAS41091.
XX
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
XX diseases
XX
XX Claim 11; SEQ ID NO 1217; 1180pp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the

| | |
|----------|---|
| CC | diagnosis, treatment, prevention and/or prognosis of a wide range of |
| CC | disorders including hyperproliferative disorders (e.g. cancer), |
| CC | immunodeficiency disorders (e.g. AIDS) autoimmune disorders |
| CC | (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), |
| CC | metabolic disorders (e.g. phenylketonuria), inflammatory disorders |
| CC | (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), |
| CC | blood-related disorders (e.g. haemophilia), reproductive disorders |
| CC | (e.g. infertility) and infectious disorders (e.g. influenza). The |
| CC | polynucleotides of the invention can also be used in gene therapy. |
| CC | AAU22915-AU233814 represent the novel human enzyme polypeptides of the |
| CC | invention. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification, but was obtained in electronic format directly from WFO |
| CC | at ftp.wipo.int/pub/published_pct_sequences . |
| SO | Sequence 146 AA: |
| Qy | Query Match 42.4%; Score 571; DB 22; Length 146; |
| Db | Best Local Similarity 97.4%; Pred. NO. 1e-51; |
| | Matches 112; Conservative 1; Mismatches 2; Indels 0; Gaps 0; |
| Qy | 1 HPETLVKVKDAEDQLGARVGYIELDLNSGELLESPFRRFPKMSIFKYLICGAVLSRID 60 |
| Db | 28 HPETLVKVKAAEDQLGARVGYIELDLNSGILLIESFRPERFPMKSTFKYLICGAVLSRID 87 |
| Qy | 61 AGCEQLGRRIHYSONDLVEYSPYTERHLLDTGNTVRELCSAATYMSDNTAANLLLT 115 |
| Db | 88 AGCEQLGRRIHYSONDLVEYSPYTERHLLDTGNTVRELCSAATYMSDNTAANLLLT 142 |
| RESULT 3 | |
| ID | AA090870 |
| XX | AA090870 standard; Protein; 146 AA. |
| AC | AA090870; |
| XX | |
| DT | 07-NOV-2001 (first entry) |
| DE | Human immune/haematopoietic antigen SEQ ID NO:18463. |
| DE | |
| KW | Human; immune; haematopoietic; immune/haematopoietic antigen; cancer; |
| KW | cytostatic; gene therapy; vaccine; metastasis. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200157182-A2. |
| XX | |
| PD | 09-AUG-2001. |
| XX | |
| PF | 17-JAN-2001; 2001WO-US01354. |
| XX | |
| PR | 31-JAN-2000; 2000US-0179065. |
| PR | 04-FEB-2000; 2000US-0180628. |
| PR | 24-FEB-2000; 2000US-0184664. |
| PR | 02-MAR-2000; 2000US-0186350. |
| PR | 16-MAR-2000; 2000US-0186374. |
| PR | 17-MAR-2000; 2000US-0190176. |
| PR | 18-APR-2000; 2000US-0198123. |
| PR | 19-MAY-2000; 2000US-0205515. |
| PR | 07-JUN-2000; 2000US-0209467. |
| PR | 28-JUN-2000; 2000US-0214886. |
| PR | 30-JUN-2000; 2000US-0215135. |
| PR | 07-JUL-2000; 2000US-0216647. |
| PR | 07-JUL-2000; 2000US-0216880. |
| PR | 11-JUL-2000; 2000US-0217487. |
| PR | 11-JUL-2000; 2000US-0217496. |
| PR | 14-JUL-2000; 2000US-0218290. |
| PR | 26-JUL-2000; 2000US-0220963. |
| PR | 26-JUL-2000; 2000US-0220964. |
| PR | 14-AUG-2000; 2000US-0224518. |
| PR | 14-AUG-2000; 2000US-0224519. |
| PR | 14-AUG-2000; 2000US-0225213. |
| PR | 14-AUG-2000; 2000US-0225214. |

| | | |
|----|-------------|----------------|
| PR | 14-AUG-2000 | 2000US-0225266 |
| PR | 14-AUG-2000 | 2000US-0225267 |
| PR | 14-AUG-2000 | 2000US-0225268 |
| PR | 14-AUG-2000 | 2000US-0225270 |
| PR | 14-AUG-2000 | 2000US-0225447 |
| PR | 14-AUG-2000 | 2000US-0225757 |
| PR | 14-AUG-2000 | 2000US-0225758 |
| PR | 14-AUG-2000 | 2000US-0225759 |
| PR | 18-AUG-2000 | 2000US-0226279 |
| PR | 22-AUG-2000 | 2000US-0226681 |
| PR | 22-AUG-2000 | 2000US-0226688 |
| PR | 23-AUG-2000 | 2000US-0227182 |
| PR | 23-AUG-2000 | 2000US-0227189 |
| PR | 30-AUG-2000 | 2000US-0228294 |
| PR | 01-SEP-2000 | 2000US-0229287 |
| PR | 01-SEP-2000 | 2000US-0229293 |
| PR | 01-SEP-2000 | 2000US-0229344 |
| PR | 01-SEP-2000 | 2000US-0229345 |
| PR | 05-SEP-2000 | 2000US-0229509 |
| PR | 05-SEP-2000 | 2000US-0229553 |
| PR | 06-SEP-2000 | 2000US-0230437 |
| PR | 06-SEP-2000 | 2000US-0230438 |
| PR | 08-SEP-2000 | 2000US-0231242 |
| PR | 08-SEP-2000 | 2000US-0231243 |
| PR | 08-SEP-2000 | 2000US-0231244 |
| PR | 08-SEP-2000 | 2000US-0231413 |
| PR | 08-SEP-2000 | 2000US-0231414 |
| PR | 08-SEP-2000 | 2000US-0232080 |
| PR | 08-SEP-2000 | 2000US-0232081 |
| PR | 12-SEP-2000 | 2000US-0232196 |
| PR | 14-SEP-2000 | 2000US-0232397 |
| PR | 14-SEP-2000 | 2000US-0232399 |
| PR | 14-SEP-2000 | 2000US-0232399 |
| PR | 14-SEP-2000 | 2000US-0232400 |
| PR | 14-SEP-2000 | 2000US-0232401 |
| PR | 14-SEP-2000 | 2000US-0233064 |
| PR | 14-SEP-2000 | 2000US-0233065 |
| PR | 14-SEP-2000 | 2000US-0233065 |
| PR | 21-SEP-2000 | 2000US-0234224 |
| PR | 21-SEP-2000 | 2000US-0234224 |
| PR | 21-SEP-2000 | 2000US-0234224 |
| PR | 27-SEP-2000 | 2000US-0235834 |
| PR | 27-SEP-2000 | 2000US-0235836 |
| PR | 29-SEP-2000 | 2000US-0236327 |
| PR | 29-SEP-2000 | 2000US-0236327 |
| PR | 29-SEP-2000 | 2000US-0236368 |
| PR | 29-SEP-2000 | 2000US-0236368 |
| PR | 29-SEP-2000 | 2000US-0236369 |
| PR | 29-SEP-2000 | 2000US-0236370 |
| PR | 02-OCT-2000 | 2000US-0236802 |
| PR | 02-OCT-2000 | 2000US-0237037 |
| PR | 02-OCT-2000 | 2000US-0237038 |
| PR | 02-OCT-2000 | 2000US-0237039 |
| PR | 02-OCT-2000 | 2000US-0237040 |
| PR | 13-OCT-2000 | 2000US-0239335 |
| PR | 13-OCT-2000 | 2000US-0239337 |
| PR | 13-OCT-2000 | 2000US-0240960 |
| PR | 20-OCT-2000 | 2000US-0241221 |
| PR | 20-OCT-2000 | 2000US-0241221 |
| PR | 20-OCT-2000 | 2000US-0241785 |
| PR | 20-OCT-2000 | 2000US-0241786 |
| PR | 20-OCT-2000 | 2000US-0241808 |
| PR | 20-OCT-2000 | 2000US-0241809 |
| PR | 20-OCT-2000 | 2000US-0241826 |
| PR | 01-NOV-2000 | 2000US-0244517 |
| PR | 08-NOV-2000 | 2000US-0244574 |
| PR | 08-NOV-2000 | 2000US-0244575 |
| PR | 08-NOV-2000 | 2000US-0246477 |
| PR | 08-NOV-2000 | 2000US-0246477 |
| PR | 08-NOV-2000 | 2000US-0246478 |
| PR | 08-NOV-2000 | 2000US-0246524 |
| PR | 08-NOV-2000 | 2000US-0246524 |

| | | |
|----|---|----------------|
| CC | 08-NOV-2000 | 2000US-0246525 |
| PR | 08-NOV-2000 | 2000US-0246526 |
| PR | 08-NOV-2000 | 2000US-0246527 |
| PR | 08-NOV-2000 | 2000US-0246528 |
| PR | 08-NOV-2000 | 2000US-0246532 |
| PR | 08-NOV-2000 | 2000US-0246609 |
| PR | 08-NOV-2000 | 2000US-0246610 |
| PR | 08-NOV-2000 | 2000US-0246611 |
| PR | 08-NOV-2000 | 2000US-0246613 |
| PR | 17-NOV-2000 | 2000US-0249207 |
| PR | 17-NOV-2000 | 2000US-0249208 |
| PR | 17-NOV-2000 | 2000US-0249209 |
| PR | 17-NOV-2000 | 2000US-0249210 |
| PR | 17-NOV-2000 | 2000US-0249211 |
| PR | 17-NOV-2000 | 2000US-0249212 |
| PR | 17-NOV-2000 | 2000US-0249213 |
| PR | 17-NOV-2000 | 2000US-0249214 |
| PR | 17-NOV-2000 | 2000US-0249215 |
| PR | 17-NOV-2000 | 2000US-0249216 |
| PR | 17-NOV-2000 | 2000US-0249217 |
| PR | 17-NOV-2000 | 2000US-0249218 |
| PR | 17-NOV-2000 | 2000US-0249244 |
| PR | 17-NOV-2000 | 2000US-0249245 |
| PR | 17-NOV-2000 | 2000US-0249264 |
| PR | 17-NOV-2000 | 2000US-0249265 |
| PR | 17-NOV-2000 | 2000US-0249297 |
| PR | 17-NOV-2000 | 2000US-0249299 |
| PR | 17-NOV-2000 | 2000US-0249300 |
| PR | 01-DEC-2000 | 2000US-0250160 |
| PR | 01-DEC-2000 | 2000US-0250391 |
| PR | 05-DEC-2000 | 2000US-0251030 |
| PR | 05-DEC-2000 | 2000US-0251988 |
| PR | 05-DEC-2000 | 2000US-0256719 |
| PR | 06-DEC-2000 | 2000US-0251479 |
| PR | 08-DEC-2000 | 2000US-0251856 |
| PR | 08-DEC-2000 | 2000US-0251859 |
| PR | 08-DEC-2000 | 2000US-0251989 |
| PR | 08-DEC-2000 | 2000US-0251990 |
| PR | 11-DEC-2000 | 2000US-0254097 |
| PR | 05-JAN-2001 | 2001US-0259678 |
| XX | | |
| PA | (HUMA-) HUMAN GENOME SCI INC. | |
| PI | Rosen CA, Barash SC, Ruben SM; | |
| XX | | |
| DR | WPI: 2001-483426/52. | |
| XX | N-PSDB: AAK63651. | |
| XX | | |
| PT | Nucleic acids encoding human immune/hematopoietic antigen polypeptides, | |
| PT | useful for preventing, diagnosing and/or treating cancers and | |
| PT | metastasis - | |
| XX | | |
| PS | Claim 11; SEQ ID NO 18463; 3071pp + Sequence Listing; English. | |
| XX | | |
| CC | AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) | |
| CC | amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic | |
| CC | activity, and can be used in gene therapy and vaccine production. (I) | |
| CC | proteins and polynucleotides may be used in the prevention, diagnosis and | |
| CC | treatment of diseases associated with inappropriate (I) expression. For | |
| CC | example, they may be used to treat disorders associated with decreased | |
| CC | expression by rectifying mutations or deletions in a patient's genome | |
| CC | that affect the activity of (I) by expressing inactive proteins or to | |
| CC | supplement the patient's own production of (I). Additionally, (I) | |
| CC | polynucleotides may be used to produce the secreted (I), by inserting the | |
| CC | nucleic acids into a host cell and culturing the cell to express the | |
| CC | protein. (I) proteins and polynucleotides may be used to prevent, | |
| CC | diagnose and treat immune/hematopoietic-related diseases, especially | |
| CC | cancers and cancer metastases of hematopoietic-derived cells. AAK64703 | |
| CC | to AAK87654 represent human immune/hematopoietic antigen genomic | |
| CC | sequences from the present invention. AAK54942 to AAK54950 and AAM82169 | |
| CC | represent sequences used in the exemplification of the present invention. | |
| XX | | |

SO Sequence 146 AA;
Query Match 42.4%; Score 571; DB 22; Length 146;
Best Local Similarity 97.4%; Pred. No. 1e-51;
Matches 112; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 HPELVKVKDAEDQAGARVGYIELDNSGELLESPREPERPPMSTFVLLCGAVLSRID 60
Db 28 HPELVKVKDAEDQAGARVGYIELDNSGELLESPREPERPPMSTFVLLCGAVLSRID 87

QY 61 AGOEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAAITMSDNTAAILLT 115
Db 88 AGOEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVRELCSAAITMSDNTAAILLT 142

RESULT 4
ID AAM20440 standard; protein; 127 AA.
XX AAM20440;
XX
XX AAM20440;
XX
DT 14-JUL-1997 (first entry)
XX
XX H. pylori cytoplasmic protein, 34574062.aa.
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; genome;
KW replication; transcription; recombination; repair.
XX
XX Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Misc-difference 121
FT /note= "encoded by codon YGC"
FT Misc-difference 122
FT /note= "encoded by codon WGC"
XX
PN WO6640893-A1.
XX
PD 19-DEC-1996.
XX
PF 06-JUN-1996; 96WO-US09122.
XX
PR 01-APR-1996; 96US-0630405.
PR 07-JUN-1995; 95US-0487032.
XX
XX (ASTR) ASTRA AB.
PI Berglindh OT, Smith D, Mellgaerd BL;
XX
XX WPI; 1997-052306/05.
DR N-PSDB; AAT67613.
XX
XX Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter
XX
XX Claim 61; Pages 615-616; 1481pp; English.
XX
XX The present sequence is a Helicobacter pylori cytoplasmic protein
CC involved in genomic replication, transcription, recombination and repair.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds, useful
CC as potential H. pylori life cycle activators or inhibitors. The genomic
CC sequence of H. pylori (ATCC 55679) was determined from overlapping
CC contigs generated by mechanically shearing the bacterial DNA. The
CC sequences were analysed for ORF of at least 180 nucleotides, and the
CC predicted coding regions defined by computer evaluation. To identify
CC likely H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be

CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.
XX
XX Sequence 127 AA;
Query Match 36.5%; Score 492; DB 18; Length 127;
Best Local Similarity 99.0%; Pred. No. 1.7e-43;
Matches 96; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPELVKVKDAEDQAGARVGYIELDNSGELLESPREPERPPMSTFVLLCGAVLSRID 60
Db 24 HPELVKVKDAEDQAGARVGYIELDNSGELLESPREPERPPMSTFVLLCGAVLSRID 83

QY 61 AGOEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVREL 97
Db 84 AGOEQLGRRIHYSQNDLVEYSPYTERKHLTDGMTVREL 120

RESULT 5
ID AAU23220 standard; protein; 129 AA.
XX AAU23220
XX
XX AAU23220;
XX
XX 18-DEC-2001 (first entry)
XX
XX
XX ~ Novel human enzyme polypeptide #306.
DE
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
XX
XX WO200155301-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.

QY 1 HPETLVKDAEDQAGARVGYIELDLNSGEILSEPRPERFPMSTPKYLICGAVLSRID 60
DB 28 HPETLVKDAEDQAGARVGYIELDLNSGKILSEPRPERFPMSTPKYLICGAVLSRID 87
QY 61 AAGEQAGRRIRHYSQNDLVEYSPYTERKHLTDGMTYARE 96
DB 88 AAGEQAGRRIRHYSQNDLVEYSPYTERKHLTDGMTYARE 123

RESULT 6
ID AAM90871 standard; Protein: 129 AA.
XX AAM90871;
AC AAM90871;
XX 07-NOV-2001 (first entry)
DE Human Immune/haematopoietic antigen SEQ ID NO:18464.
XX Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX Homo sapiens.
OS Homo sapiens.
PN WO200157182-A2.
XX 09-AUG-2001.
PD 17-JAN-2001; 2001WO-US01354.
PE 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.

```
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249289.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR MPI: 2001-483426/52.
XX
PT N-PSDB; AAK63652.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
PS Claim 11: SEQ ID NO 18464; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 129 AA;
Query Match 35.4%; Score 477; DB 22; Length 129;
Best Local Similarity 96.9%; Pred. No. 6.5e-42;
Matches 93; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 HPETLVKVKDAEDLGARVGYIELDINSGETLSPRPERFRPMSTFKVLLCGAVLSRID 60
DB 28 HPETLVKVKDAEDLGARVGYIELDINSKILSPRPERFRPMSTFKVLLCGAVLSRID 87
OY 61 AGQEQGLRRIRIHYSQNDLVEYSPTVEKHLTDGMTVRE 96
DB 88 AGQEQGLRRIRIHYSQNDLVEYSPTVEKHLTDGMTXME 123
RESULT 7
ABG27917
ID ABG27917 standard; Protein; 182 AA.
AC ABG27917;
```

```
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #27908.
DE
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HISE-) HISEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX MPI: 2001-639362/73.
XX
XX N-PSDB; AAS92104.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20: SEQ ID NO 58276; 103pp; English.
XX
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 182 AA;
Query Match 35.4%; Score 477; DB 22; Length 182;
Best Local Similarity 87.3%; Pred. No. 1.1e-41;
Matches 96; Conservative 4; Mismatches 6; Indels 4; Gaps 1;
OY 1 HPETLVKVKDAEDLGARVGYIELDINSGETLSPRPERFRPMSTFKVLLCGAVLSRID 60
DB 24 HPETLVKVKDAEDLGARVGYIELDINSKILSPRPERFRPMSTFKVLLCGAVLSRID 83
OY 61 AGQEQGLRRIRIHYSQNDLVEYSPTVEKHLTDGMTVRELSAATMSDNTAA 110
DB 84 AGQEQGLRRIRIHYSQNDLVEYSPTVEKHLTDGTER----LEKSFSDNTAS 129
RESULT 8
AAB59052
ID AAB59052 standard; Protein; 94 AA.
```


DB 140 VVGXIG 145

RESULT 10
ABG27935

ID ABG27935 standard; Protein: 101 AA.

XX ABG27935;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #27926.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS92122.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 58294; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 101 AA;

SQ Query Match 27.6%; Score 372; DB 22; Length 101;

Best Local Similarity 100.0%; Pred. No. 4.9e-31;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 DTFNPVAMATLRLKLLTGELLTLASRQQLDWMWADKVPAPLLRSALPAGFIADKSGAG 213
|||||
DB 10 DTFNPVAMATLRLKLLTGELLTLASRQQLDWMWADKVPAPLLRSALPAGFIADKSGAG 69
|||||

QY 214 ERGSRGIIAALGPD 227
|||||
DB 70 ERGSRGIIAALGPD 83RESULT 11
ABG27919

ID ABG27919 standard; Protein: 88 AA.

XX ABG27919;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #27910.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS92106.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 58278; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations in
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 88 AA;

SQ Query Match 24.7%; Score 333; DB 22; Length 88;

Best Local Similarity 81.2%; Pred. No. 5e-27;

Matches 69; Conservative 3; Mismatches 3; Indels 10; Gaps 1;

QY 4 TLVAVKDAEQLGKRVGYIELDNGEILSESPERPFPMSTFKYLCAVLSRIDAGQ 63
|||||

RESULT 14
AA045138 standard; Protein; 158 AA.
ID AA045138
AC AA045138;
DT 27-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #6034.
DE
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
PN MO200181581-A2.
PD
XX
XX 01-NOV-2001.
PF 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'Malsonneuve J, Zhang Y, Jen S, Carter D;
XX
XX MPI: 2001-616774/71.
DR N-PSDB; AAS59525.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1; SEQ ID No 6333; 1069pp; English.
XX
XX Sequences AA039105-AA068017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 158 AA;
Query Match 5.9%; Score 79.5; DB 22; Length 158;
Best Local Similarity 33.3%; Pred. No. 4.9;
Matches 26; Conservative 9; Mismatches 36; Indels 7; Gaps 2;

Db 135 GRMGSWLATRIPLRW 152
| : ||: | || |
RESULT 15
AAW55524
ID AAW55524 standard; Protein; 200 AA.
AC AAW55524;
XX
XX 02-JUL-1998 (first entry)
DT
XX
XX H. pylori ORF 29ep10720_24432762_c3_39 cellular protein.
DE
XX
XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KM identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
XX
XX Helicobacter pylori.
OS
XX
XX MO937044-A1.
PN
XX
XX 09-OCT-1997.
PD
XX
XX 27-MAR-1997; 97WO-US05223.
PF
XX
XX 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
XX (ASTR) ASTRA AB.
XX
XX Alm RA, Smith D;
PI
XX
XX MPI: 1997-503122/46.
DR N-PSDB; AAW24933.
XX
XX Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
XX Claims 14, 93; Page 726-727; 1145pp; English.
XX
XX This sequence is a H. pylori cellular protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds. The
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions for
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SQ Sequence 200 AA;
Query Match 5.9%; Score 79; DB 18; Length 200;
Best Local Similarity 26.0%; Pred. No. 7.8;
Matches 51; Conservative 31; Mismatches 44; Indels 70; Gaps 13;

Thu Sep 11, 16:07:50 2003

seq2_30e.rag

Page 13

```

Db      11 VVLDTDSOKSMETATITAEKERPTFSLF-----NNSGGFSDTLKQMSKYENIL 61
OY      79 ----EXSPVTEKHLTDGMTVRELCSAITSNDNTAANLL--TIGGPKEL-TAFILHG 131
        || || ||
Db      62 DMGEFSKEKOK-----AMLLS-----NIVLPTF--PSOLDEVLANN 99
OY      132 DHTVRLDRWEPELNE--AIPNDERDTMPVAMATLRLKLLTGELTFLASROQLDME- 187
        : : : || | | : | : |
Db      100 ERLEQLO----ELWENLRALIVINRPTIP-----TLKERALTEFIKE 139
OY      188 ---ADKVAGPLLRSL 200
        : : : || | |
Db      140 NNPSDRIT--LLESSL 153

```

```
Search completed: September 10, 2003, 12:24:50
Job time : 33.7143 secs
```

THIS PAGE BLANK (USPTO)

NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-991-890-4

Query Match 5.3%; Score 71; DB 3; Length 159;
Best Local Similarity 23.8%; Pred. No. 6.3;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

QY 49 VLICGAVLSRIDAGE-----QLGRIRHSQN-----DLVEYSPYTERK 87
DB 12 LILCGAYF--VSPQEIHAERFGRHRRHHHGGSGAELRGCGPRFGKHLLSYCPMPERT 69
QY 88 LTDCMTRELCSAITSNDTANLLTTIGPKELTAFLHMGDHYTRLDREPELNEA 147
DB 70 FT-----TPGGWLLS--GRKENVSTSNKDGQAL-----GTTSEF 105
QY 148 IPNDERDTMPVAMA--TLRLKL 169
DB 106 IPNLSPELKKPLSSEQPSLKKTI 128

RESULT 5
US-09-252-991A-24241
Sequence 24241, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24241
LENGTH: 197
TYPE: PRN
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match 5.2%; Score 70.5; DB 4; Length 197;
Best Local Similarity 23.6%; Pred. No. 10;
Matches 37; Conservative 22; Mismatches 43; Indels 55; Gaps 7;

QY 94 VRELCSAITSNDTANLLTTIGPKELTAFLHMGDHYTRLDREPELNEAIPNDER 153
DB 12 VAKISSARTSASASLASVPLIPATP--TSARFSAGSLT-----PSPTVI 55
QY 154 DTMPPVA--MATTLKRLTGLTLLA-----SRQLDWMEDAKVAGP----- 194
DB 56 ATSPSLACRACRTRSLCSG--LARANTSTPGSTSRRR--WSSSNSISAPVAGRSMP 110
QY 195 -----ILRSALPAGFIADKSGAGRG 216
DB 111 SMAPLARAFTWSTWPTVITTAIPAAMHSAATAASSRG 147

RESULT 6
US-09-010-809-19
Sequence 19, Application US/09010809B
Patent No. 6090601
GENERAL INFORMATION:
APPLICANT: Gustafsson, Claes
TITLE OF INVENTION: Ectothione polypeptide synthases and encoding DNA
FILE REFERENCE: 30062-20020.00
CURRENT APPLICATION NUMBER: US/09/010,809B
CURRENT FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 158
TYPE: PRN
ORGANISM: Sorangium cellulosum
US-09-010-809-19

Query Match 4.9%; Score 65.5; DB 3; Length 158;
Best Local Similarity 27.6%; Pred. No. 25;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

QY 106 DNTANLLT-----TIGPKELTAFLHMGDHYT--RLDRWEP-----ELNEAIPNDER 154
DB 21 NHDAKHLITSRQASADGADVLRSELALGASVTLACDVPALRLKLDLNDIP----- 75
QY 155 TTMVVAAMATLRLKLTGLTLLASRQQLDWMEDAKVAGLNSA-----LPAGWF 205
DB 76 SAHPVAANVHAASVDGDLGAMSLERT-----DRVFAKIDAAWHLHDTODKPLAAF 129
QY 206 IADKSGAGRGSRG 219
DB 130 ILFSSVAGVGLSSG 143

RESULT 7
US-09-252-991A-29942
Sequence 29942, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29942
LENGTH: 178
TYPE: PRN
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (116)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match 4.9%; Score 65.5; DB 4; Length 178;
Best Local Similarity 26.8%; Pred. No. 30;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

QY 110 ANLLITIG---GKELTAFL---LHMGDHYTRLDREPELNEAIPNDERDTMPVAMA 162
DB 25 ARYLVLVLAQORRRGELAAHAGLAAHHPGDHVAR-----AGEDLP---RDVQJLAAA 74
QY 163 TTLRLKLTGLL-----TLASRQQLDWMEDAKVAGLNSALPAGWFIADKSG 211

Db 75 LRHLHYLAGALOGHEHHEGIGDGTAGEQAVVG--ODQEVAPXV--GLQAGLFLWQGD 130
QY 212 AGERGSGIIALGPDK 229
Db 131 A-----LVGVGQAGK 141

RESULT 8

US-09-413-814-87
Sequence 87, Application US/09413814
Patent No. 6225064

GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofte, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
TITLE OF INVENTION: heteropolypeptide compounds
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 87
LENGTH: 198
TYPE: PRT

ORGANISM: Sorangium cellulosum
US-09-413-814-87

Query Match 4.8%; Score 64.5; DB 3; Length 198;
Best Local Similarity 22.4%; Pred. No. 46;
Matches 62; Conservative 30; Mismatches 78; Indels 107; Gaps 15;

QY 1 HPETLVKVAADQAGARVGYTLDLNGSELSEFRPEERPMNSTF-----KVLGCAV 55
Db 7 NPEAVDAVDKTS--AAFGV---DATVHKVLEGGCAQVETAIYTFGEHFEPSYLCVAS 60
QY 56 LSRIDAGQEQGLRRIRIYSONDLVYFSPYTEKHLTDGATVRELCSAATMSDNTAANLLT 115
Db 61 LVR-----MGVRIIARA---TDRQAD--ILR 82
QY 116 TIGPKRELTAFLHNMGDHYTRLDRWEPELNEAIPNDERDTTPVAMATTLRLKILGELLT 175
Db 83 AVG-----ATRIQLLETMGRRV---GADITMPLAQ-----DILD 114
QY 176 LASRQQLDMEADKVAQPLRLSALPAGWFIADK---SGAGER-----GSGRIIALG 225
Db 115 LASHRRVVPW---NAHGFLVGTGTL-AGSKIRQRYRINVLGVRPTNRKPGKRLFAET 169
QY 226 PDGKPSRIVVIYTTGQATM-----DERNRQIAEIG 256
Db 170 PD-----YVRDGDTLTLVGSDSDVSGRFAEVG 197

RESULT 9

US-09-252-991A-24921
Sequence 24921, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24921
LENGTH: 203
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24921

Query Match 4.8%; Score 64.5; DB 4; Length 203;
Best Local Similarity 27.1%; Pred. No. 48;
Matches 35; Conservative 16; Mismatches 41; Indels 37; Gaps 7;

QY 135 TRLDWRPEL-----NEAIPNDERDTTPVAMATTLRLKILGELLTSLASRQQLDME 187
Db 34 TPLDSESLRLRVSLAMWRKMRPRPMQARRPMKLAET--RALTGISTLSSR-----DWS- 86
QY 188 ADKVAQPLRLSALPAGWFIADKSGAGERSGRIIALGPDKPSRIVVIYTTGQATMDE 247
Db 87 -----TSAMPFGCDI-----SPGMLASTRLBCK-AKLATATAT---AALKV 123
QY 248 RNRQIAEIG 256
Db 124 PTRYSTIG 132

RESULT 10

US-09-239-909-2
Sequence 2, Application US/09239909
Patent No. 6284952

GENERAL INFORMATION:
APPLICANT: Kumbo Petrochemical Co. Ltd.
TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
FILE REFERENCE: P99-2-6
CURRENT APPLICATION NUMBER: US/09/239,909
CURRENT FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: EP 99300136.1
EARLIER FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: KOPATIN 1.0
SEQ ID NO 2
LENGTH: 150
TYPE: PRT
ORGANISM: G. max calmodulin4 (SCAM4)
US-09-239-909-2

Query Match 4.7%; Score 64; DB 3; Length 150;
Best Local Similarity 27.9%; Pred. No. 34;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

QY 73 SONDLVEXSPYTEKHLTDG---MTVRELCSAATMSDNTAANLLTTI-----GGPRE 122
Db 6 SEQIVDFREARGLFPGKDDGCTIVELATVIRLDQPTREELQDMSEVDADNGTIE 65
QY 123 LTAFLHNMGDHYTRLDRWEPELNEAIPNDERDTTPVAMATTLRLKL--TGELLTSLASRQ 180
Db 66 FDEFSLMAKKVQDTDA--EELKEAFKVFQDKDGYIS--ASELRVIMINIGSEKIDEEVE 123
QY 181 QLIDWMEAD 189
Db 124 QMT--KEAD 130

RESULT 11

US-08-624-677A-2
Sequence 2, Application US/08624677A
Patent No. 6476192

GENERAL INFORMATION:

APPLICANT: Tally, Nicola C.
APPLICANT: Jenkins, Mark C.


```

Query Match      4.7%: Score 63; DB 4; Length 174;
Best Local Similarity 23.9%; Pred. No. 56;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4

OY          7 KYKADBDOLGAR-VGYIELDINSGBLESFREFRFPNMTSTFKVLLCGAVLSRIDAGCEQ 65
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB          67 KKEGATTELKKNIPILVKVDCTEEBALCRDGVGEGLPIIKIFGL-----DAVKPY 117
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY          66 LGRS-----IHYS-QNDLVESPTYTEKHITDGMVRELCSSAITSMDNTANILLTT 116
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB          94 OGAROTEAIVSYMWKSLPAVSPTPENLEETKTMKDVIYIGIYASDDGTANDIEFTT 174
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14
US-08-557-122A-12
Sequence 12, Application US/08557122A
Patent No. 5879664

GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSER: NO. 5879664dsk of NO. 5879664th America, Inc.
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,122A
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lambdis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3980, 204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEX: 212-878-9655
INFORMATION FOR SEO ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 200 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-122A-12

Query Match      4.7%: Score 63; DB 2; Length 200;
Best Local Similarity 23.9%; Pred. No. 69;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4,

OY          7 KYKADBDOLGAR-VGYIELDINSGBLESFREFRFPNMTSTFKVLLCGAVLSRIDAGCEQ 65
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB          43 KKEGATTELKKNIPILVKVDCTEEBALCRDGVGEGLPIIKIFGL-----DAVKPY 93
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY          66 LGRS-----IHYS-QNDLVESPTYTEKHITDGMVRELCSSAITSMDNTANILLTT 116
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB          94 OGAROTEAIVSYMWKSLPAVSPTPENLEETKTMKDVIYIGIYASDDGTANDIEFTT 150
            | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
US-09-262-666-12
Sequence 12, Application US/09262666
Patent No. 6346244

GENERAL INFORMATION:
APPLICANT: Hjort, Carsten Malland
TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
NUMBER OF SEQUENCES: 38

```

```

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: NO. 63462440 No. 6346244disk of No. 6346244th America, Inc.
3 STREET: 405 Lexington Avenue, 64th Floor
4 CITY: New York
5 STATE: New York
6 COUNTRY: United States of America
7 ZIP: 10174-6401
8 COMPUTER READABLE FORM:
9 MEDIDM TYPE: Floppy disk
10 COMPUTER: IBM PC compatible
11 OPERATING SYSTEM: PC-DOS/MS-DOS
12 SOFTWARE: PatentIn Release #1.0, Version #1.30
13 CURRENT APPLICATION DATA:
14 APPLICATION NUMBER: US/09/262,666
15 FILING DATE:
16 CLASSIFICATION:
17 PRIOR APPLICATION DATA:
18 APPLICATION NUMBER: US 08/557,122
19 FILING DATE: 11-DEC-1995
20 ATTORNEY/AGENT INFORMATION:
21 NAME: Lambiris, Elias J.
22 REGISTRATION NUMBER: 33,728
23 REFERENCE/DOCKET NUMBER: 3980.204-US
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE: 212-867-0123
26 TELEFAX: 212-878-9655
27 INFORMATION FOR SEQ ID NO: 12:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 200 amino acids
30 TYPE: amino acid
31 TOPOLOGY: linear
32 MOLECULE TYPE: protein
33 US-09-262-666-12
34
35 Query Match 4.7%; Score 63; DB 4; Length 200;
36 Best Local Similarity 23.9%; Pred. No. 69;
37 Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4.
38
39 7 KYVDADDOIGAR-VGYIFELDINSGLILESFRREERPPMSTFVVLGCAVLSRIDAQEO 65
40 I::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
41 43 KTEQDAATELKEKNIPLVKVDCTEEBALCRDQGEVGYPTLKIFRGL-----DAVKPY 93
42
43 66 LGRR-----IHY5-QNDLVERSPVTEKHILDTMTVRELCSAATMSDNTANILTTT 116
44 :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
45 94 QCARQTEALVSYNWKSLPAVSPYTPENLEETKMDKIIVIGIAYISDQTDANDIFTT 150

```

Search completed: September 10, 2003, 12:31:02
Job time : 12.8571 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds

(without alignments)
2183,941 Million cell updates/sec

Title: SEQ2_30E

Perfect score: 1348
Sequence: 1 HPEITLVKVKADADICARVG.....TMDERNQIAICASLIRHW 263

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCN_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCNUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|---------------------|---------------------|
| 1 | 463 | 34.3 | 94 | US-10-102-806-760 | Sequence 760, App |
| 2 | 78 | 5.8 | 202 | US-10-127-816-9 | Sequence 9, App1 |
| 3 | 78 | 5.8 | 202 | US-10-142-717-12 | Sequence 12, App1 |
| 4 | 69.5 | 5.2 | 149 | US-10-233-926-4 | Sequence 4, App1 |
| 5 | 69 | 5.1 | 149 | US-10-156-761-8136 | Sequence 8136, App |
| 6 | 69 | 5.1 | 178 | US-09-738-626-5680 | Sequence 5680, App |
| 7 | 69 | 5.1 | 206 | US-10-156-761-13867 | Sequence 13867, App |
| 8 | 67.5 | 5.0 | 184 | US-10-101-464A-765 | Sequence 765, App |
| 9 | 67.5 | 5.0 | 187 | US-09-948-018-8 | Sequence 8, App1 |
| 10 | 67 | 5.0 | 194 | US-10-156-761-8623 | Sequence 8623, App |
| 11 | 67 | 5.0 | 202 | US-10-127-816-11 | Sequence 11, App1 |
| 12 | 65.5 | 4.9 | 191 | US-09-764-870-390 | Sequence 390, App |
| 13 | 65.5 | 4.9 | 191 | US-10-125-540-390 | Sequence 757, App |
| 14 | 65 | 4.8 | 179 | US-09-764-868-757 | Sequence 4858, App |
| 15 | 65 | 4.8 | 179 | US-10-106-698-4858 | |

| | | | | | |
|----|------|-----|-----|---------------------|---------------------|
| 16 | 65 | 4.8 | 202 | US-10-189-346-12 | Sequence 12, App1 |
| 17 | 64.5 | 4.8 | 191 | US-10-156-761-8978 | Sequence 8978, App |
| 18 | 64.5 | 4.8 | 193 | US-09-805-354-8 | Sequence 8, App1 |
| 19 | 64.5 | 4.8 | 193 | US-10-144-259-8 | Sequence 8, App1 |
| 20 | 63.5 | 4.7 | 189 | US-09-815-242-5862 | Sequence 5862, App |
| 21 | 63.5 | 4.7 | 189 | US-09-815-242-12979 | Sequence 12979, App |
| 22 | 63.5 | 4.7 | 206 | US-09-738-626-5425 | Sequence 5425, App |
| 23 | 62.5 | 4.6 | 144 | US-10-131-406-4 | Sequence 4, App1 |
| 24 | 62.5 | 4.6 | 190 | US-10-156-761-9507 | Sequence 9507, App |
| 25 | 62.5 | 4.6 | 195 | US-10-156-761-12656 | Sequence 12656, App |
| 26 | 62 | 4.6 | 162 | US-09-738-626-4796 | Sequence 4796, App |
| 27 | 62 | 4.6 | 174 | US-09-864-761-35777 | Sequence 35777, App |
| 28 | 62 | 4.6 | 174 | US-09-864-761-35777 | Sequence 4637, App |
| 29 | 62 | 4.6 | 202 | US-10-738-626-4637 | Sequence 4637, App |
| 30 | 61.5 | 4.6 | 188 | US-10-100-252-6 | Sequence 6, App1 |
| 31 | 61.5 | 4.6 | 142 | US-09-862-027-31 | Sequence 31, App1 |
| 32 | 61.5 | 4.6 | 187 | US-10-156-761-12111 | Sequence 12111, App |
| 33 | 61.5 | 4.6 | 195 | US-10-156-761-6624 | Sequence 6624, App |
| 34 | 61.5 | 4.6 | 202 | US-10-189-346-16 | Sequence 16, App1 |
| 35 | 61 | 4.5 | 70 | US-09-864-761-41555 | Sequence 41555, App |
| 36 | 61 | 4.5 | 177 | US-09-791-933-69 | Sequence 69, App1 |
| 37 | 61 | 4.5 | 203 | US-09-800-729-154 | Sequence 154, App1 |
| 38 | 60.5 | 4.5 | 152 | US-10-156-761-14346 | Sequence 14346, App |
| 39 | 60.5 | 4.5 | 160 | US-09-882-227-230 | Sequence 230, App |
| 40 | 60.5 | 4.5 | 196 | US-10-219-220-265 | Sequence 265, App |
| 41 | 60.5 | 4.5 | 200 | US-10-156-761-12086 | Sequence 12086, App |
| 42 | 60 | 4.5 | 134 | US-09-768-2358-40 | Sequence 40, App1 |
| 43 | 60 | 4.5 | 184 | US-09-798-029-4 | Sequence 4, App1 |
| 44 | 60 | 4.5 | 184 | US-09-795-926-46 | Sequence 46, App1 |
| 45 | 60 | 4.5 | 184 | US-10-364-774-46 | Sequence 46, App1 |

ALIGNMENTS

RESULT 1
US-10-102-806-760
? Sequence 760, Application US/10102806
? Publication No. US20030054421A1
? GENERAL INFORMATION:
? APPLICANT: Rosen et al.
? TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
? FILE REFERENCE: PA103P1C1
? CURRENT APPLICATION NUMBER: US/10/102,806
? PRIOR FILING DATE: 2002-03-22
? PRIOR APPLICATION NUMBER: 09/925,298
? PRIOR FILING DATE: 2001-08-10
? PRIOR APPLICATION NUMBER: PCT/US00/05881
? PRIOR FILING DATE: 2000-03-08
? PRIOR APPLICATION NUMBER: 60/124,270
? PRIOR FILING DATE: 1999-03-12
? NUMBER OF SEQ ID NOS: 846
? SOFTWARE: Patent Ver. 2.0
? SEQ ID NO 760
? LENGTH: 94
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (80)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION: (91)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760

Query Match 34.3% Score 463; DB 15; Length 94;
Best Local Similarity 98.9%; Pred. No. 2.9e-40;

Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 75 NDLEVSPTVEKHLTDGKTVRELCSSAATMSDNTANLLTTIGSPKEITAFIHNNGDHV 134
DB 1 NDLEVSPTVEKHLTDGKTVRELCSSAATMSDNTANLLTTIGSPKEITAFIHNNGDHV 60

OY 135 TRLDWEPELNEAIPNDEROTMPVAMATT 164
DB 61 TRLDWEPELNEAIPNDERXTMPVAMATT 90

RESULT 2

US-10-127-816-9
; Sequence 9, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Kluehner, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127, 816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-9

Query Match 5.8%; Score 78; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 6.2; Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEQELGRRIHYSQND--LVEYSPVTEKHL-----TDGMTVREL---- 97
DB 17 LLLAAVLTQTQADVPVPRATRLPVEAKDCHIAQFSLSPKELQAFKAKADIEKRLLEKDL 76
OY 98 -CSA-----AATMS--DNTAANLLTTTGGPKRELTAFLHN 129
DB 77 RCSSHLFPRAWDKQLOVQERPKALQAEVALTLTKWEMNTDSALATILGQPLHTLSIHS 136
OY 130 MGDHVT-----RLDRWEPELNEAIPNDEROTMPVAMATTLLTGELTL 176
DB 137 QLQCTQLQATAEPSPSRRLSRWLRLQEA-QSKETPCCLASVTSNLFRLTLTDLKCV 195
OY 177 ASRQQLI 183
DB 196 ANGDCV 202

RESULT 3

US-10-142-717-12
; Sequence 12, Application US/10142717
; Publication No. US20030104579A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Baum, Peter R.
; APPLICANT: Mosley, Bruce A.
; APPLICANT: Ketchum, Randal R.
; APPLICANT: Taylor, Scott L.
; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
; CURRENT FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-142-717-12

Query Match 5.8%; Score 78; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 6.2; Matches 40; Conservative 27; Mismatches 66; Indels 54; Gaps 7;

OY 50 LCGAVLSRIDAGEQELGRRIHYSQND--LVEYSPVTEKHL-----TDGMTVREL---- 97
DB 17 LLLAAVLTQTQADVPVPRATRLPVEAKDCHIAQFSLSPKELQAFKAKADIEKRLLEKDL 76
OY 98 -CSA-----AATMS--DNTAANLLTTTGGPKRELTAFLHN 129
DB 77 RCSSHLFPRAWDKQLOVQERPKALQAEVALTLTKWEMNTDSALATILGQPLHTLSIHS 136
OY 130 MGDHVT-----RLDRWEPELNEAIPNDEROTMPVAMATTLLTGELTL 176
DB 137 QLQCTQLQATAEPSPSRRLSRWLRLQEA-QSKETPCCLASVTSNLFRLTLTDLKCV 195
OY 177 ASRQQLI 183
DB 196 ANGDCV 202

RESULT 4

US-10-233-926-4
; Sequence 4, Application US/10233926
; Publication No. US20030131382A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
; FILE REFERENCE: BB1419 US NA
; CURRENT APPLICATION NUMBER: US/10/233,926
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: US/09/735,846
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
US-10-233-926-4

Query Match 5.2%; Score 69.5; DB 16; Length 149;

Best Local Similarity 20.8%; Pred. No. 31; Matches 22; Conservative 25; Mismatches 34; Indels 25; Gaps 4;

OY 156 TMPVANATTLRLKIL-----TGELTLASROQLIDWMEADKVAQPLRSALPAGNFIADK 209
DB 11 SLSLSLSKSLPNIILAAADHAAEAAPGSSQDEEDMKAEAGGDGV-----EYADR 61
OY 210 SGAGERSRGIIAALGPDGKPSRIV-----IYTTGSOATYDERRN 250
DB 62 GGGCGAANGSI-----PEGRRIRYADSIYDLFHHGAKSLSQANR 102

RESULT 5

US-10-156-761-8136
; Sequence 8136, Application US/10156761
; Publication No. US20030119018A1

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8136
; LENGTH: 149
; TYPE: PR1
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8136
```

```

Query Match
Best Local Similarity 5.1%; Score 69; DB 15; Length 149;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;
```

```

QY 168 LITTELT-----ASRQQLDMEDRVNAGPLRSALPAGWFIADKSG----- 211
Db 50 LLDIELTILKRLVAVSVKAKEMGIDMWEHD---PALSHADGGRLAEENRLREI 105
QY 212 AGERGRGIIALGDKRSR 232
Db 106 AGLRDAQALPSAEGPAEKPER 126
```

```

RESULT 6
US-09-738-626-5680
; Sequence 5680, Application US/0978626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIKOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHITO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5680
```

```

Query Match
Best Local Similarity 5.1%; Score 69; DB 10; Length 178;
Matches 34; Conservative 19; Mismatches 65; Indels 36; Gaps 5;
```

```

QY 113 LITTCPEKELTAFLEHMGD--VTRLDREWEPELENEAIPNDERPTMPVAMATRLKLT 170
Db 27 LKVTAGKRSNAIAIKVDSRPLDQLEVSQETGELFDAAEQKELNFGAGYTLVEVSTP 86
QY 171 GELLTLASRQQLIDWNEADRVNAGPLRSALPAGWFIADKSGAGERSGIIALGPDGR 230
Db 87 G-----VDNPL---TLPRH-----RRNGRLVALDQDGK 114
QY 231 --SRIVYITTSQATMDERNRQIAETGASLKH 262
Db 115 RVARIGALMDAETHVLIERNKLEVTLELAH 148
```

```

RESULT 7
US-10-156-761-13867
; Sequence 13867, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13867
; LENGTH: 206
; TYPE: PR1
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13867
```

```

Query Match
Best Local Similarity 5.1%; Score 69; DB 15; Length 206;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11;
```

```

QY 70 IHVSQNDLVEVSPVEKHLTDGTV-----RELCSAATMGDNATLITIT---G 118
Db 1 VHASOG---AGPGTESAGEEYKTMADAGESEFEYVA-----NRSALTKTVLISG 50
QY 119 GPK-----LTFALHNMGDHVRILD-----RW-----EPELENAIPNDE 152
Db 51 GDRHAEDLLQNALIRAADRWSRIDEPEAVYRVLYRQVSRWRLKWRRELVAEP-PE 109
QY 153 RDTTPVAMATRLKLTGELLTLASRQQL-----DWEADRVNAGPLRSALPAGW 205
Db 110 ASTGPDASAABELRVNRLGALRLTANQFVLVRYFEDLPDAD--VARITIGCSVGVRS 167
QY 206 IADKSGAGERSGIIALGPD---DGKPSR 232
Db 168 TTHRSIARLRLTAPELAAALGPADAEQEPSR 197
```

```

RESULT 8
US-10-101-464A-765
; Sequence 765, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
```

```
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 765
; LENGTH: 184
; TYPE: PRN
; ORGANISM: Pinus radiata
US-10-101-464A-765
```

```
Query Match
Best Local Similarity 24.7%; Pred. No. 67;
Matches 43; Conservative 22; Mismatches 64; Indels 45; Gaps 9;
```

```
OY 41 FPMSTFKYL-----LCGAVLSRIDAGQEQGLRITHYSONDLVESPV---TEKHLTGD 91
DB 23 FCMSTYERLLVYPYMLNGSVASRL-----RDSINGKPALDMPTRKRISIG 67
OY 92 -----MTVRELCSAATMSDNTANLL-----TTIGREKELTAFLHMGDHTRLDRM-- 140
DB 68 AARGLLYLHQCPKRIHRVYKANNILDEYFEAVVCGFLAKLLDHRSDHVTAVRGTV 127
OY 141 ---EPLENEAIPDERDTPVNAATTLRIKLTGEL---LTIASRQ--LIDGM 186
DB 128 GHAPPYLSTGGSSEK--TDVFGGILLLELTIGOKALDFGRANOKGVMLDW 179
```

```
RESULT 9
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: The11 et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 8
; LENGTH: 187
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-948-018-8
```

```
Query Match
Best Local Similarity 22.4%; Pred. No. 68;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;
```

```
OY 112 LLLTITGGREKELTAFLHMGDHTRLDRME-----LNEAIPDERDTPVNAAT- 163
DB 11 LLLALAPPEAS-----QYCGRLTYWNPDKNCSSCLQRFGR-----PPCPGALFTG 58
OY 164 -----TLRKLTLGCELITLASR--QOLIDMWEA-----DKVAGP-----L 195
DB 59 DTWKASLLPLLSRELSSASQPLSRILDELVEVEELIYVLDPEPGCGGMAGTTRHLLA 118
OY 196 LRSALPAGWFIADKSGAGSGRGIITIALGPDGKPSR--IYVITTSQATDERRRQIA 253
DB 119 AHYGLPAAW-----STFAYSLRPSRPLRLIENVARREPSASIGQLGTHLLA 165
OY 254 EIG 256
DB 166 QIG 168
```

```
RESULT 10
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
```

```
; APPLICANT: OKURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKARI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
```

```
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 8623
; LENGTH: 194
; TYPE: PRN
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623
```

```
Query Match
Best Local Similarity 38.6%; Pred. No. 81;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2;
```

```
OY 202 AGWEIADKSGAGGERSRG---ITAAIGPDG--KPSRIYIYTT 239
DB 75 SGFRVTPRSGAGERVTEGAEYRTTAIGPLAVHEVRVAVVVT 118
```

```
RESULT 11
US-10-127-816-11
; Sequence 11, Application US/10127816
; Publication No. US20030104416A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Rindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 11
; LENGTH: 202
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-127-816-11
```

```
Query Match
Best Local Similarity 5.0%; Score 67; DB 15; Length 202;
```



```
RESULT 15
US-10-106-698-4858
: Sequence 4858, Application US/10106698
: Publication No. US20030109690A1
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
: FILE REFERENCE: PA005P1
: CURRENT APPLICATION NUMBER: US/10/106,698
: CURRENT FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: Patent In Ver. 3.0
: SEQ ID NO 4858
: LENGTH: 179
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-106-698-4858

Query Match          4.88; Score 65; DB 15; Length 179;
Best Local Similarity 27.38; Pred. No. 1.2e+02;
Matches 24; Conservative 18; Mismatches 34; Indels 12; Gaps 4;

OY      120  PKELTAFLHMGDHYTRLD--RWPELNEA-----IPNDRDTMPVAMATTIRKLTG 171
          |||||  |::  ||  |||  ||::  |  |  |
DB      52  PRELTVV---QGKLEVLDHSHKRWMLYKNENAGRGCTIPSNILEPLQPGCTPGTQCGSPSRV 108
          |||||  ::|||  ::||::  ::||  |  |  |
OY      172  ELLTLASR-QQLIDNMEADRVAGPLRS 198
          |||||  ::|||  ::||::  ::||  |  |  |
DB      109  PMLRLSSRPEVTDWTLQANENSTATVRT 136
          |||||  ::|||  ::||::  ::||  |  |  |
```

Search completed: September 10, 2003, 12:33:17
Job time : 17.5714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds

(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E
Perfect score: 1348
Sequence: 1 HPETLVKKAEDQLGARVY.....TMDERNRQIAIGASLIKHW 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|-------------|
| 1 | 524 | 38.9 | 105 | 2 | JC2566 |
| 2 | 78 | 5.8 | 191 | 2 | S67447 |
| 3 | 75.5 | 5.6 | 190 | 2 | T09136 |
| 4 | 75.5 | 5.6 | 200 | 2 | G97064 |
| 5 | 74 | 5.5 | 177 | 2 | A83753 |
| 6 | 73.5 | 5.5 | 195 | 2 | T36975 |
| 7 | 72.5 | 5.4 | 113 | 2 | T45195 |
| 8 | 71.5 | 5.3 | 152 | 2 | T36984 |
| 9 | 71 | 5.3 | 192 | 2 | G83096 |
| 10 | 71 | 5.3 | 198 | 2 | C68296 |
| 11 | 70.5 | 5.2 | 131 | 2 | AD2281 |
| 12 | 70.5 | 5.2 | 184 | 2 | T21126 |
| 13 | 70 | 5.2 | 108 | 2 | E84217 |
| 14 | 69.5 | 5.2 | 145 | 2 | F84251 |
| 15 | 68.5 | 5.1 | 167 | 2 | D87360 |
| 16 | 68.5 | 5.1 | 172 | 2 | AD3606 |
| 17 | 68.5 | 5.1 | 180 | 2 | C71869 |
| 18 | 68 | 5.0 | 170 | 2 | AB0192 |
| 19 | 67.5 | 5.0 | 192 | 2 | A83587 |
| 20 | 67.5 | 5.0 | 198 | 2 | D95285 |
| 21 | 66.5 | 4.9 | 128 | 2 | F87353 |
| 22 | 66 | 4.9 | 149 | 2 | F71252 |
| 23 | 65.5 | 4.9 | 42 | 2 | F56978 |
| 24 | 65.5 | 4.9 | 116 | 2 | C82906 |
| 25 | 65.5 | 4.9 | 195 | 2 | AE0623 |
| 26 | 65 | 4.8 | 146 | 2 | C72703 |
| 27 | 65 | 4.8 | 148 | 2 | E75283 |
| 28 | 65 | 4.8 | 150 | 2 | AC0284 |
| 29 | 65 | 4.8 | 153 | 2 | E81708 |

| | | | | | | |
|----|------|-----|-----|---|--------|--------------------|
| 30 | 65 | 4.8 | 160 | 1 | E69186 | conserved hypotnet |
| 31 | 65 | 4.8 | 168 | 2 | B75498 | conserved hypotnet |
| 32 | 65 | 4.8 | 180 | 2 | G70912 | hypothetical prote |
| 33 | 64.5 | 4.8 | 151 | 2 | D81333 | hypothetical prote |
| 34 | 64.5 | 4.8 | 177 | 2 | D90227 | probable protein-t |
| 35 | 64 | 4.7 | 150 | 2 | D70580 | hypothetical prote |
| 36 | 64 | 4.7 | 152 | 2 | T08585 | hypothetical prote |
| 37 | 64 | 4.7 | 162 | 2 | AG0769 | calmodulin - soya |
| 38 | 63.5 | 4.7 | 168 | 2 | T20606 | probable acetyltra |
| 39 | 63.5 | 4.7 | 177 | 2 | S11602 | hypothetical prote |
| 40 | 63.5 | 4.7 | 179 | 2 | AB1994 | hypothetical prote |
| 41 | 63.5 | 4.7 | 182 | 2 | A96202 | hypothetical prote |
| 42 | 63.5 | 4.7 | 196 | 2 | H69647 | nwsb protein (limp |
| 43 | 63 | 4.7 | 122 | 2 | AC3609 | 2-dehydro-3-deoxy- |
| 44 | 63 | 4.7 | 124 | 2 | F75373 | transcription regu |
| 45 | 63 | 4.7 | 168 | 2 | AG2257 | conserved hypotnet |
| | | | | | | hypothetical prote |

ALIGNMENTS

RESULT 1

JC2566 bla protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.

Gene 148[128], 81-86, 1994

A:Title: Construction of improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <MES>

C:Genetics:

A:Gene: bla

C:Superfamily: beta-lactamase I

Query Match 38.9% Score 524: DB 2: Length 105:

Best Local Similarity 100.0% Pred. No. 3.4e-37;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| QY | 160 | AMATTLRLKLTGELLTASROOLIDWMEADKVGAPLRLSALPACGFTADKSGAGERSRG | 219 |
| DB | 2 | AMATTLRLKLTGELLTASROOLIDWMEADKVGAPLRLSALPACGFTADKSGAGERSRG | 61 |
| QY | 220 | IIAALGPDGKPSRIYVYTTGSOATMDERNRQIAIGASLIKHW | 263 |
| DB | 62 | IIAALGPDGKPSRIYVYTTGSOATMDERNRQIAIGASLIKHW | 105 |

RESULT 2

S67447 hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, March 1996

A:Reference number: 221766

A:Accession: T38062

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:Z69944; NID:g1217974; PIDN:CAA93808.1; PID:g1217978; GSPDB:G

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.8% Score 78; DB 2; Length 191;

Best Local Similarity 22.6%; Pred. No. 29;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;
QY 104 MSDNTAANLLTTTGGCKELTAFIHNMGDHYTRDREPELNEALPNDERDTPYAMAT 163
DB 1 MSYMSLNLMLQPSGDKIAILVN---VARLD-----PASAKSTQVLVSMLN 46
QY 164 TRKLLTGLLTLASROQLDMWEADKVGPLLRSLAPAGWFIADKSGAGERSRGITAA 223
DB 47 EPRC-----ILRLPGLYKLIYNFRKDSPEPYMSNAINIGYVTE--GLAFLCGKQITSI 99
QY 224 LQPDGKP-----SRIVVYITGTSQATNDERNQI----- 252
DB 100 ----SKPLEDKMLMSSRFMLDTLITVYQLREKTEDEKREQDLASMLASLPICITMS 155
QY 253 AEIGASLILKH 262
DB 156 VENGAGLHKH 165

RESULT 3
T09136
ADP-ribosylation factor homolog ARL3 - Trypanosoma brucei
N:Alternate names: ADP ribosylation factor 3 homolog
C:Species: Trypanosoma brucei
C:Date: 11-Jun-1999 #sequence, revision 11-Jun-1999 #text_change 02-Feb-2001
C:Accession: T09136
R:Bringingaud, F.; Vedrene, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.;
MOL. Biochem. Parasitol. 94, 249-264, 1998
A:Title: Conserved organization of genes in trypanosomatids.
A:Reference number: Z16580; MUID:98418771; PMID:9747975
A:Accession: T09136
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190 <BRI>
A:Cross-references: EMBL:AR031926; NID:93452211; PIDN:AAC32774.1; PID:93452215
A:Experimental source: strain Antali
A>Note: small G-protein
C:Genetics:
A:Gene: ARL3
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)
F:89-94/Region: nucleotide-binding motif B
F:126-129/Region: GTP-binding NKXD motif
F:171-176/Region: nucleotide-binding motif B
F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

Query Match 5.6%; Score 75.5; DB 2; Length 190;
Best Local Similarity 20.5%; Pred. No. 46;
Matches 46; Conservative 37; Mismatches 78; Indels 63; Gaps 9;
QY 30 ELLESRRPERFRPMSTFKVILCGAVLSRIDAGQQLGRIRIHYSONDLVEYSPTERKHLT 89
DB 5 EFLKLRPPSR---RTRRLMLGL---DNAGKTRLLRRICEE-----VSDTFPT 48
QY 90 DQMTVEELCSAATMSNTAANLLTTTGGCKELTAFIHNMGD-----VTRL 137
DB 49 QCFNIONITADEL-----KPYVMDVGQKSLRSYMRHYDPDHDALFVIDSAMERI 100
QY 138 DWPEPLNEALPNDERDTPYAMATTLRKLLTGLLTLASROQLDMWEADKVGPL-L 196
DB 101 EEARTEHLYLEEE-----KLIVGVPDLLFANKQDIPKASQOEVMSSLNL 145
QY 197 RSALPAGWFIADKSG-AGERSRGII-----AALGPDGK 229
DB 146 ADTINRPMHIELCSAETGEGLSLSWVDTLKKRPSLRPDCQ 189

RESULT 4
G97064
spore coat protein COTJC [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence, revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.
J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97064
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79306.1; PID:Q15024270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1338

Query Match 5.6%; Score 75.5; DB 2; Length 200;
Best Local Similarity 25.7%; Pred. No. 50;
Matches 36; Conservative 26; Mismatches 55; Indels 23; Gaps 7;
QY 1 HPEVLVKKDAEDQLGARVGYIELDLSNGELIESFR-PEEREPMSTFKVILCGAVLSRI 59
DB 10 HP---VKIKMPNQL-AKVIITQYGGPDGELAASIRYLSQRSWTP-----QATATLN 59
QY 60 DAGQQLGRIRIHYSONDLVEYSPTERKHLTDGVTRELCSAIT--MSDNTAANLLTTI 117
DB 60 DIGTEELAH-----LEIVGSIVQLSGLSVEELKSGLDAYFADHDSAIYPASAA 110
QY 118 GGPKELTAFIHNMGDHYTRL 137
DB 111 GNP-PTAAVYIQSKGDPITDL 129

RESULT 5
A83753
hypothetical protein BH0825 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence, revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A83753
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83753
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:q10173176; PIDN:BA804544.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0825
C:Superfamily: Bacillus subtilis hypothetical protein ykka

Query Match 5.5%; Score 74; DB 2; Length 177;
Best Local Similarity 20.3%; Pred. No. 56;
Matches 38; Conservative 29; Mismatches 56; Indels 64; Gaps 8;
QY 42 PMWSTFKVILCGAVLSRIDAGQQLGRIRIHYSONDLVEYSPTERKHLTDGVTRELCSSA 101
DB 11 PEMDTSVGLFAYAVEENVND-----RLH-----HLIEDVTEELYYRG 47
QY 102 ITMSDNTAANLLTTTGGCKELTAFIHNMGCHVTRLD-RWPEL-NEALPND----- 151
DB 48 SDGDSNSMQLT-----NHLTFVDVWRVFRIGEGALPDSLEAEGPM 89
QY 152 -ERDTPYAMATTLRKLLTGLLTLASRQ-----LDMWEADKVGAVPLLSALP 201
DB 90 VDKGRLPVYTSVQELIEKORYVALALKETCOALHDDDLARWIPYEEROATIRWGL- 148
QY 202 AGWFIAD 208
DB 149 --WHMAD 153

RESULT 6
T36975
hypothetical protein SCJ11.04 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36975
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: 221618
A:Accession: T36975
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-195 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CAB52889.1; GSPDB:GN00070; SCOEDB:SCJ11.04
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.04

Query Match 5.5%; Score 73.5; DB 2; Length 195;
Best Local Similarity 17.0%; Pred. No. 71;
Matches 40; Conservative 43; Mismatches 77; Indels 75; Gaps 9;
QY 15 LGARVYIEL-DLNGSEILSESPEREPFPMSTFKVLLCGAVLSRIDAGOEOLGRHYS 73
DB 17 MASMGLLEARASAREVEVLRRE-----AARAVALENGEIELDRV--- 60
QY 74 ONDLVEYSPVTEKHITDGMTVRELCSAATMSDNTAANLLTTGGPRELAFILNMGDH 133
DB 61 -----IAEEIEVALVSAEFTGTAESEGETALVPAPASAEPCAI----- 104
QY 134 VTRLDREPELNEAI--PNDERTTMPVAMATILKRLTGLTLASROQLIDMEADKY 191
DB 105 ---VPHWEGLSVSVSPNNOR-----ILNVLDDRPGLPEPARAKDI 142
QY 192 AGPLRLSALPA-----GMFIADKSGAGERSRGITIAALGPDGKPS 231
DB 143 AAALGIEAAAVAEVGRPKGRKRLAEKGLQLQEASGASAGNR-LVAS--PGGDPS 194

RESULT 7
T45195
hypothetical protein U1756t [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: 216911
A:Accession: T45195
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

Query Match 5.4%; Score 72.5; DB 2; Length 113;
Best Local Similarity 28.6%; Pred. No. 41;
Matches 34; Conservative 18; Mismatches 44; Indels 23; Gaps 8;
QY 130 MGDHV--TRLDREPELN--EAIPIIDERDT---TMPVAMATILKRLTGLTLASRO 180
DB 1 MGDGIGMEREGRMTGTQCPRLRVPGDESEPTLDGRASPEDLIIT--NLSPTIMSHPPS 58
QY 181 QLIIDMMEA-DKVAAGPL-----RSALDPAGWFIADKSGAGERSRGITIAALGPDGKPS 232
DB 59 RDDDMVEEPFDALOGTAVFADATGDKATMPAVGGI---GASTRGS-GILASLSPFROPAR 112

RESULT 8
T36984
hypothetical protein SCJ11.13 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000

C:Accession: T36984
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: 221618
A:Accession: T36984
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-152 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CAB52889.1; GSPDB:GN00070; SCOEDB:SCJ11.13
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCJ11.13
C:Superfamily: Streptomyces coelicolor hypothetical protein SCJ9A.17

Query Match 5.3%; Score 71.5; DB 2; Length 152;
Best Local Similarity 23.4%; Pred. No. 74;
Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;
QY 106 DNTAANLLTTGGPRELAFILNMGDHYTRLDREPELNEAIPNDERDTMPVAMATTL 165
DB 25 DQAAARTLAQVTPPTPADLATFYESIGD-VN---WEDVGNQYFLNP----- 65
QY 166 RKLITGELLTLASROQLIDMEADKYAGPLRLSALPAGWFIADKSGAGERSRGITIAALG 225
DB 66 ---AGDILLRLQEGVDFGTDKSRGLVI-----GSKG-GLITYAG 104
QY 226 PDGKPSRIVYITGSOATMDERN-----ROIARI 255
DB 105 PDG-----AVYRTRT-ASLDEAEKLVADGIRQLFL 135

RESULT 9
G83096
conserved hypothetical protein PA4399 [imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001
C:Accession: G83096
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
A:Reference number: A82950; MIMD:20437337; PMID:10984043
A:Accession: G83096
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-192 <STO>
A:Cross-references: GB:AE004855; GB:AE004091; MID:G9950621; PIDN:AA07787.1; GSPDB:GN
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA4399
C:Superfamily: conserved hypothetical protein DRI638

Query Match 5.3%; Score 71; DB 2; Length 192;
Best Local Similarity 26.0%; Pred. No. 11e+02;
Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;
QY 53 GAVLSRI-----DAGOEQL--GRRIHSONDLVEYSPTEKHITDGMTVRELCSA----- 100
DB 2 GNRLSKIYTRTDGRTGAGRRVPKSHPRIEAIGAVDELNSQLGLLALTELRGAP 61
QY 101 -----ATMSDNTAANLLTTGGPRELAFILNMGDHYTRLDREPEL-----N 145
DB 62 GLEIIVQALAPVOHR-----LFDLGEELAPFEYRAIDETEVARLESCIDRNWDELGP 116
QY 146 EAIPIIDERDTMPVAMATILKRLTGLTLASROQLIDMEADKYAGPLRLSALPAGW 205
DB 117 FILPGSR---PVAQAHVCRLAR-----SAERRCOLDDEETLEGVGLRYLNLRLDL 168
QY 206 IADKSGAGERSRGIT--AALGPD 227
DB 169 VAARAIARQGVAEILMEAAAKPD 192

RESULT 10

C69296

conserved hypothetical protein AF0371 - *Archaeoglobus fulgidus*C:Species: *Archaeoglobus fulgidus*

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 29-Sep-1999

C:Accession: C69296

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.

; Glodek, A.; Zhou, L.; Overbeek, R.; Cooney, J.D.; Weisman, J.F.; McDonald, L.

Nature 350, 364-370, 1997

A:Authors: Ultebach, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.

Smith, H.O.; Moose, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: C69296

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-198 <MLE>

A:Cross-references: GB:AE001079; GB:AE000782; NID:g2689402; PIDN:AAB90876.1; PID:g265027

C:Superfamily: hypothetical protein M0724

Query Match 5.3%; Score 71; DB 2; Length 198;

Best Local Similarity 23.6%; Pred. No. 1.2e+02;

Matches 50; Conservative 26; Mismatches 80; Indels 56; Gaps 12;

QY 14 QLCARVGYELDLNNGEILSPREPERPMSTFKVLLCGAVL---SRIDAG---QEQLG 67

DB 15 QLCER-----GEVEFDFRFLDFSVATIRTELAFCISTANSSATAGLQFRLG 64

QY 68 RHINYSQ-----NDLVEYSPVTEKHITDGMTVELCSAATMSNTANLLTTI 117

DB 65 QCVGVEALTLACVRFHNRKAEY-----FRAKSPKLVKALAEASSKAEILKIK 117

QY 118 G-GPKELTFLANNG-DHATRLD---RMEDELNEAIPNDERDTMPVAMATLTKLL-- 169

DB 118 GIGMKASHLRLVGRDVAIIDRIHLEROGYEVC-----TMTAKYLEVEKILME 172

QY 170 ----TGEILLTASRQQLIDMMEADKAGPLLR 197

DB 173 ISEERGESL---AEMDLRIWAE---MTGCKVLK 198

RESULT 11

AD2281

hypothetical protein alr3803 [imported] - *Nostoc* sp. (strain PCC 7120)C:Species: *Nostoc* sp. PCC 7120A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AD2281

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuch

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete genome sequence of the filamentous Nitrogen-fixing Cyanobacterium *Ana*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AD2281

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-131 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075502.1; PID:g17132937; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3803

Query Match 5.2%; Score 70.5; DB 2; Length 131;

Best Local Similarity 26.0%; Pred. No. 73;

Matches 34; Conservative 19; Mismatches 51; Indels 27; Gaps 5;

QY 77 LVEYSPVTEKHITDGMTVELCSAA-ITMSDNTANLLTTIGPKELTAPLHNM--GDH 133

DB 1 MYSSSVTLQNIQNVLFLLFSSONCOLYDIQTEPLDPPYYSKEFNRRTALVAGNF 60

QY 134 VTRLD--RMEPELNEAIPNDERDTMPVAMATLTKLLT-----GE 172

DB 61 VLSFDWTNMQDEANRFYTNPE---LNLNADISTICKLLTSVHKERFCSGHLAQMDNGH 117

QY 173 LTLTASRQQLI 183

DB 118 LTLTQRLQAI 128

RESULT 12

T21126

ADP-ribosylation factor homolog F19H8.3 [similarity] - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001

C:Accession: T21126

R:Steward, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: T21126

A:Accession: T21126

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-184 <WIL>

A:Cross-references: EMBL:293378; PIDN:CA07583.1; GSPDB:GN00020; CESP:F19H8.3

A:Experimental source: clone F19H8

C:Genetics:

A:Gene: CESP:F19H8.3

A:Map position: 2

A:Introns: 143/2

C:Superfamily: ADP-ribosylation factor

C:Keywords: blocked amino end; lipoprotein; myristylation; nucleotide binding; P-loop

F:24-31/Region: nucleotide-binding motif A (P-loop)

F:90-95/Region: nucleotide-binding motif B

F:127-130/Region: GTP-binding NKXD motif

F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

Query Match 5.2%; Score 70.5; DB 2; Length 184;

Best Local Similarity 21.0%; Pred. No. 1.2e+02;

Matches 45; Conservative 40; Mismatches 80; Indels 49; Gaps 11;

QY 30 EILSPREPERPMSTFKVLLCGAVLSRIDAGQOLRRIRHYSQNDLVEYSPVTEKHIT 89

DB 5 DVLKSRK-----PSGRIRIRILLGL-----DNAGKRTIILKQL--SSDVGHPV-----T 48

QY 90 DGMTVELCSAATMSDNTANLLTTIGPKELTAPLHNMGNH-----TRL 137

DB 49 KGFNVKTVAA-----MCD---IRLNVMDIGGGRSIRPYSNYYENIDTLFVIDSNKKRF 101

QY 138 DKMEPELNEAIPNDERDTMPVAMATLTKLLTGLTLTASRQQLIDMMEADKAGPLLR 197

DB 102 DENNIRIGELT-DEEKLRKVPVLIFANKQ-----DLVTAASSSEITRKINLTD-----LLR 150

QY 198 SALPAGMFIADKSGAGRGSRGIIAALGPDRKPS 231

DB 151 DRT--WHIQACSAALKNEGINDTITWASNLKPA 181

RESULT 13

E84217

hypothetical protein vng0594h [imported] - *Halobacterium* sp. NRC-1C:Species: *Halobacterium* sp. NRC-1

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: E84217

R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergulst, B.; Pan, M.; Shukla, H.D.; Lasky

; Leitner, B.; Keller, K.; Cruz, R.; Danon, M.J.; Hough, D.W.; Maddocks, D.G.; Ja

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of *Halobacterium* species NRC-1.

A:Reference number: AB4160; MUID:20504483; PMID:11016950

A:Accession: E84217

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-108 <STO>

A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AG19105.1; GSPDB:GN00138

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds

(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E
Perfect score: 1348
Sequence: 1 HPETLVKVKDAEDQLGARVGT.....TMDERNRQIAIEIGASLIKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 78 | 5.8 | 191 | 1 | YDA4_SCHPO |
| 2 | 71 | 5.3 | 198 | 1 | OGG1_ARCFU |
| 3 | 69.5 | 5.2 | 196 | 1 | RAC2_LOTJA |
| 4 | 69 | 5.1 | 201 | 1 | RACG_DICDI |
| 5 | 67 | 5.0 | 122 | 1 | HGJL_THIFE |
| 6 | 66 | 4.9 | 149 | 1 | NDR_TREPA |
| 7 | 65.5 | 4.9 | 116 | 1 | RBRF_UREPA |
| 8 | 65.5 | 4.9 | 206 | 1 | KTHY_METAC |
| 9 | 63.5 | 4.7 | 196 | 1 | ALRH_BACST |
| 10 | 63.5 | 4.7 | 200 | 1 | RR4_PELNE |
| 11 | 63 | 4.7 | 176 | 1 | HSIV_THEMA |
| 12 | 62.5 | 4.6 | 144 | 1 | MANR_ECOLI |
| 13 | 62.5 | 4.6 | 172 | 1 | YDEJ_ECOLI |
| 14 | 62.5 | 4.6 | 182 | 1 | PYRE_STRCO |
| 15 | 62.5 | 4.6 | 195 | 1 | TRPF_THEVO |
| 16 | 62 | 4.6 | 174 | 1 | THB_RAT |
| 17 | 62 | 4.6 | 182 | 1 | YB61_BIFLO |
| 18 | 61.5 | 4.6 | 184 | 1 | MUR1_SCHPO |
| 19 | 61.5 | 4.6 | 185 | 1 | YCJC_ECOLI |
| 20 | 61.5 | 4.6 | 197 | 1 | HANI_XENLA |
| 21 | 61 | 4.5 | 177 | 1 | VNSC_RINKA |
| 22 | 61 | 4.5 | 177 | 1 | VNSC_RINDR |
| 23 | 61 | 4.5 | 202 | 1 | COAT_ELV |
| 24 | 60.5 | 4.5 | 126 | 1 | YGM1_YEAST |
| 25 | 60.5 | 4.5 | 142 | 1 | YB63_METTM |
| 26 | 60.5 | 4.5 | 146 | 1 | HBG_PABIT |
| 27 | 60.5 | 4.5 | 150 | 1 | PDUV_SALTY |
| 28 | 60.5 | 4.5 | 160 | 1 | YATB_HELPY |
| 29 | 60.5 | 4.5 | 175 | 1 | Y581_AOUAF |
| 30 | 60.5 | 4.5 | 178 | 1 | HSIV_RALSO |
| 31 | 60 | 4.5 | 121 | 1 | SECR_HUMAN |
| 32 | 60 | 4.5 | 159 | 1 | GREX_CHUTE |
| 33 | 60 | 4.5 | 178 | 1 | PYRE_ARCFU |

| | | | | | | |
|----|------|-----|-----|---|------------|--------------------|
| 34 | 60 | 4.5 | 178 | 1 | UCR1_ANASP | P70758 anabaena sp |
| 35 | 60 | 4.5 | 184 | 1 | ARL2_DROME | Q06849 drosophila |
| 36 | 60 | 4.5 | 200 | 1 | TATB_CAUCR | Q06849 drosophila |
| 37 | 59.5 | 4.4 | 103 | 1 | RS10_NEIGO | P48851 neisseria g |
| 38 | 59.5 | 4.4 | 152 | 1 | YU33_YERPE | Q08268 yersinia pe |
| 39 | 59.5 | 4.4 | 156 | 1 | BFR_AZOVI | P22759 azotobacter |
| 40 | 59.5 | 4.4 | 177 | 1 | ATPD_HAEIN | P43717 haemophilus |
| 41 | 59.5 | 4.4 | 184 | 1 | HRPL_PSESY | P37929 pseudomonas |
| 42 | 59.5 | 4.4 | 198 | 1 | UPP_ARATH | O65583 arabidopsis |
| 43 | 59.5 | 4.4 | 205 | 1 | RS4_ECOLI | P02354 escherichia |
| 44 | 59.5 | 4.4 | 206 | 1 | RS4_SHIFL | P59132 shigella fl |
| 45 | 59.5 | 4.4 | 206 | 1 | RS4_YERPE | Q08288 yersinia pe |

ALIGNMENTS

RESULT 1

| ID | YDA4_SCHPO | STANDARD | PRT | 191 AA. |
|----|--|----------|-----|---------|
| AC | Q10346; | | | |
| DT | 01-OCT-1996 (Rel. 34, Created) | | | |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) | | | |
| DT | 28-FEB-2003 (Rel. 41, Last annotation update) | | | |
| DE | Hypothetical protein C1F12.04c in chromosome I. | | | |
| GN | SPAC1F12.04C. | | | |
| OS | Schizosaccharomyces pombe (fission yeast). | | | |
| OC | Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; | | | |
| OC | Schizosaccharomycetales; Schizosaccharomycetaceae; | | | |
| OC | Schizosaccharomycetes. | | | |
| OX | NCBI_TaxID:4896; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN:972; | | | |
| RX | MEDLINE-21848401; PubMed-11859360; | | | |
| RA | Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., | | | |
| RA | Sgouras K., Peat N., Hayles J., Baker S., Basham D., Bowman S., | | | |
| RA | Brown D., Brown S., Chillingworth T., Churcher C.M., | | | |
| RA | Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A., | | | |
| RA | Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., | | | |
| RA | Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., | | | |
| RA | James K., Jones L., Jones M., Leather S., McDonald S., McLean J., | | | |
| RA | Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., | | | |
| RA | Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E., | | | |
| RA | Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., | | | |
| RA | Skellon J., Slimmons M., Squares R., Squares S., Stevens K., | | | |
| RA | Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S., | | | |
| RA | Woodward J., Voicakert G., Aert R., Robben J., Gilmour B., | | | |
| RA | Wells J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., | | | |
| RA | Gabel C., Fuchs M., Filtz C., Holzer E., Moestl D., Hilbert H., | | | |
| RA | Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M., | | | |
| RA | Berger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., | | | |
| RA | Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., | | | |
| RA | Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., | | | |
| RA | Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G., | | | |
| RA | Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., | | | |
| RA | Domiguez A., Revuelta J.L., Moreno S., Armstrong J., Forzburg S.L., | | | |
| RA | Cerritelli L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., | | | |
| RA | Shpakovski G.V., Ussery D., Barrett B.G., Nurse P., | | | |
| KT | The genome sequence of Schizosaccharomyces pombe." | | | |
| RL | Nature 415:871-880(2002). | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (see http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | EMBL: 269944; CAA93808.1; - | | | |
| DR | PIR: T38062; S67447. | | | |
| DR | GenBank: SPAC1F12.04C; - | | | |

QY
19 VGSITLDDNGSELLESTFRPEERFPMMSTFKVLLCGAVLSRIDAGEQLGRRIHYSQNOLY 78
|| : || | || ||
| : | |
12 VGVEGDLDSXSLLEYFR-----NAILARRGGG-----LG 41
Db

Db 42 KYAPFTGTYEIGVT-----ISGDKDPQAGLVLT---PREST 75

RESULT 6

NCB_TREPA STANDARD; PRT; 149 AA.
AC 083974;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
DE (Nucleoside-2-P kinase).
GN NDK OR TP1010.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBITaxID=160;

RP SEQUENCE FROM N.A.
RC STRAIN-Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Welstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete.
RL Science 281:375-388(1998).
CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
CC OTHER THAN ATP.
CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate - ADP +
CC nucleoside triphosphate.
CC -1- SUBUNIT: Homotrimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the NDK family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: AE001269; AAC65961.1; -.
DR HAMAP: MF_00451; -; 1.
DR HAMAP: MF_00451; -; 1.
DR InterPro: IPR001564; NDK.
DR Pfam: PF00334; NDK; 1.
DR PRINTS: PR01243; NUCDPKINASE.
DR ProDom: PD01018; NDK; 1.
DR SMART: SM00562; NDK; 1.
DR PROSITE: PS00469; NDP_KINASES; 1.
KW Transferrase; Kinase; ATP-binding; Complete proteome.
FT ACT_SITE 117 117 BY SIMILARITY.
SQ SEQUENCE 149 AA; 16657 MW; BB39539671E80BE CRC64;

Query Match 4.9%; Score 66; DB 1; Length 149;
Best Local Similarity 21.6%; Pred. No. 1.3e+02;
Matches 35; Conservative 24; Mismatches 65; Indels 38; Gaps 7;

QY 29 GEIESRPERFERPMSTFKVLLCGAVLSRIDAGQDGLGRHHSQNDIVKSPYTEKHL 88
DB 21 GEVLSRF--ERKGVLTALRLC-----VDPADEL-----HAEHREKPFYSLAYI 67
QY TDGVTVRELCSAITMSDNTANLLTTIGPKRETAFLINMGDHTVLRMEPELNEAI 148
DB 68 TSAFV-----ALAFKGENAISLVTLGCS-----TRVENAQP---GTI 103

QY 149 PND-EBDTPMPVAMATTLKRLTGLTTLASRQQLDMEND 189
DB 104 RGDFAIRTTITIVHSDSPESAAREALYFSAODEFVEMRDGN 145

RESULT 7

RBFA_UREPA STANDARD; PRT; 116 AA.
AC Q9PQH0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR UV321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBITaxID=134821;

RP SEQUENCE FROM N.A.
RC STRAIN-Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glas J.I., Lefkowitz E.J., Glas J.S., Helner C.R., Chen E.Y.,
RA Cassell G.H.;
RT The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum.
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polyosomes).
CC Essential for efficient processing of 16S rRNA (By similarity).
CC the 5' terminal helix region of 16S rRNA (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: AE001229; AAF30730.1; -.
DR HAMAP: MF_00003; -; 1.
DR InterPro: IPR000238; Rib_bind_factA.
DR Pfam: PF02033; RBFA; 1.
DR ProDom: PD007327; Rib_bind_factA; 1.
DR TIGRFAMs: TIGR00082; tDfa; 1.
DR PROSITE: PS01319; RBFA; FALSE_NEG.
KW rRNA processing; Complete proteome.
SQ SEQUENCE 116 AA; 13247 MW; A79DC71F0547514 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 116;
Best Local Similarity 28.9%; Pred. No. 1.1e+02;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

QY 144 LNEAIPRDEBDTTPVAMATTLKRLTGLT-----LTLASRQQLDMME-ADRVAEPPL 196
DB 18 INNALANEINDKIARVAVNR--LSNDLSVAKIFLDHAKRSMKLVLENVNVSG--LL 74

QY 197 RSALPAGM-----FIADKS 210
DB 75 RSKLAERTSYKPELRVIDET 97

RESULT 8

KTHY_METAC STANDARD; PRT; 206 AA.
AC Q8THS9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).

GN TMK OR MA4433.
 OS Methanosarcina acetivorans.
 OC Archaea: Euryarchaeota: Euryarchaeota orders Incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 NCBI_TaxID=2214;
 RM [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., DeArliano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umeyan L.A., White O., White R.H., de Macario E.C.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Zander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity.";
 RL Genome Res. 13:532-542(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate -> ADP + thymidine
 5'-diphosphate.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: AE011164; AM07774.1; -
 DR HAMAP: MF_00165; -1-
 DR InterPro: IPR000062; Thymidylate_kin.
 DR Pfam: PF02223; Thymidylate_kin; 1.
 DR TIGRPFAM: TIGR00041; DTM_Kinase; 1.
 DR TRANSFER: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
 KW transferase; kinase; Nucleotide biosynthesis; ATP-binding;
 FT Complete proteome.
 FT NP_BIND 10
 SQ SEQUENCE 206 AA; 23303 MW; A19C2E3BC85423B3 CRC64;
 ATP (POTENTIAL).
 Query Match 4.9%; Score 65.5; DB 1; Length 206;
 Best Local Similarity 18.7%; Pred. No. 2.2e+02;
 Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;
 QY 56 LSRID-AGEOLGRIRIHVSQNDLVEYSPYTEKHLTDGMTVRELCSAATMSDNTAANLL 114
 DB 8 LEGIDSGKSTVAKRIQ-ENSELRVPEVPTREPTRG-----TLGNMVENMIQ 55
 QY 115 TTIGGPRELTAFLHMGCHVTRLDREPELENAIPNDEBRTTMEVAMATTAKLITGELL 174
 DB 56 SDTQALAEFLPTAAHAEHLATLVKPALEDGKTIVSDSRAYOGITFLKNRLNDPL 114
 QY 175 TLASROQLIDMWEADKAVGPIRLRSALPAGM-----FLND-----KSG-AGE----- 214
 DB 115 -----EVR-----DIHRGWTYIPDLTFLFDIEPELAVKRCGRGGEQTKF 154
 QY 215 -----RSGRIIAGPDGKSRIVYITTGSOATMDE 247
 DB 155 EKIEFLRGVRELFLGLAAV-EPERVIYIVASGSPREYER 192
 RESULT 9
 ALH_BACSU STANDARD; PRT; 196 AA.
 AC P50846;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE KMG/KDPG aldolase [includes: 4-hydroxy-2-oxoglutarate aldolase
 DE (EC 4.1.3.16) (2-keto-4-hydroxyglutarate aldolase) (KMG-aldolase); 2-
 DE dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14) (phospho-2-
 DE dehydro-3-deoxygluconate aldolase) (Phospho-2-keto-3-deoxygluconate
 DE aldolase)].
 DE KDG.
 GN Bacillus subtilis.
 OS Bacteria: Firmicutes: Bacillales: Bacillaceae; Bacillus.
 OC NCBI_TaxID=1423;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / Marburg;
 RX MEDLINE=96349105; PubMed=8760912;
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,
 RA Serror P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sera and kdg loci cloned in a yeast artificial chromosome.";
 RL Microbiology 143:2005-2016(1996).
 RM [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolofo A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codan J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denlioz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enlil K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grand G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Nodack M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Portoullis S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadele Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weissenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 CC -1- CATALYTIC ACTIVITY: 4-hydroxy-2-oxoglutarate -> pyruvate +
 CC glyoxylate.
 CC -1- CATALYTIC ACTIVITY: 2-dehydro-3-deoxy-D-gluconate 6-phosphate ->
 CC pyruvate + D-glyceraldehyde 3-phosphate.
 CC -1- PATHWAY: KEY ENZYME IN THE ENTNER-DOIDOROFF PATHWAY.
 CC -1- PATHWAY: PARTICIPATES IN THE REGULATION OF THE INTRACELLULAR LEVEL
 CC OF GYOXALATE.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE KMG/KDPG ALDOLASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----

```

CC -----
DR EMBL: L47836; AAB38480.1; -.
DR EMBL: 299115; CAB1427.1; -.
DR PIR: H69647; H69647.
DR Subtilisin; BG11396; kDa.
DR InterPro: IPR000887; AldIse_KDPG_KHG.
DR Pfam: PF01081; Aldolase; 1.
DR TIGRfams: TIGR01182; ead; 1.
DR PROSITE: PS00159; ALDOLASE_KDPG_KHG_1; FALSE_NEG.
DR PROSITE: PS00160; ALDOLASE_KDPG_KHG_2; 1.
DR Lyase; Schiff Base; Multifunctional enzyme; Complete proteome.
KW ACT_SITE 43 43 BY SIMILARITY.
FT ACT_SITE 47 47
FT BINDING 130 130 SCHIFF-BASE WITH KHG OR PYRUVATE (BY
FT SIMILARITY).
SQ SEQUENCE 196 AA; 20865 MW; E80B597FD1102CCA CRC64;
Query Match 4.7%; Score 63.5; DB 1; Length 196;
Best Local Similarity 21.6%; Pred. No. 3e+02;
Matches 27; Conservative 16; Mismatches 61; Indels 19; Gaps 4;
OY 28 SGELISEFRPEERPPMSTFRVYLICGAVLSRIDAQ-EOLGRRIHSON--DLVERSPV 83
DB 51 ASDIIESFNRREDI-----LIGAGTVISAOACEAKAGAFIVSPGSADLAEHLSF 103
OY 84 TEKHLTDGMIVRELCSAIFMSDTPAANL-----LTITGGPRKELTFALHMGDHYT 135
DB 104 VTHYIIPGVLPSEIMELATFGFTLLKIPSGVCSIPFKKNLAGPPQVTFIPGSLHPS 163
OY 136 RLDRW 140
DB 164 EYPDW 168
RESULT 10
RR4_PELINE STANDARD; PRT; 200 AA.
AC Q9FSD9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 30S ribosomal protein S4.
GN RPS4.
OS Pellia necosiana (Liverwort).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Jungermanniopsida; Metzgeriidae; Fossombroniales; Pellineae;
OC Pellineaceae; Pellia.
OX NCBI_TaxID=70144;
OX 11
RN RP
RP SEQUENCE FROM N.A.
RC TISSUE=gametophyte;
RA Capselin I.; Bloecher R.;
RT "A molecular approach to bryophyte systematics."
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: One of the primary rRNA binding proteins, it binds
CC directly to 16S rRNA where it nucleates assembly of the body of
CC the 30S subunit (By similarity).
CC -! FUNCTION: With S5 and S12 plays an important role in translational
CC accuracy (By similarity).
CC -! SUBUNIT: Part of the 30S ribosomal subunit. Contacts protein S5.
CC The interaction surface between S4 and S5 is involved in control
CC of translational fidelity (By similarity).
CC -! SUBCELLULAR LOCATION: Chloroplast.
CC -! SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -! SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC -----

```

```

CC      or send an email to license@lsb-sib.ch).
CC      -----
CC      EMBL: AJ250456; CAC14050.1; -.
CC      DR HSSP; P81288; 1C05.
CC      DR HAMAP; ME_01306; -: 1.
CC      DR InterPro; IPR001912; Ribosomal_S4.
CC      DR InterPro; IPR002942; S4.
CC      DR InterPro; IPR005709; S4_bact.org.
CC      DR Pfam; PF00163; Ribosomal_S4; 1.
CC      DR Pfam; PF01479; S4; 1.
CC      DR SMART; SM00363; S4; 1.
CC      DR TIGRfams; TIGR01017; tpsd_bact; 1.
CC      DR PROSITE; PS00632; RIBOSOMAL_S4; 1.
CC      DR PROSITE; PS00889; S4; 1.
CC      KW Ribosomal protein; RNA-binding; rRNA-binding; Chloroplast.
CC      FT DOMAIN 90 150 S4 RNA-BINDING.
CC      SQ SEQUENCE 200 AA; 23178 MW; 00D0098C6D933F5 CR664;

Query Match 4.7%; Score 63.5; DB 1; Length 200;
Best Local Similarity 23.4%; Pred. No. 3; 1e+02;
Matches 40; Conservative 30; Mismatches 78; Indels 23; Gaps 8;

QY 14 QUGARVGYIE-LDNGSEGLIESRPEREPFMSTFKVLGCAVLSRIDAGQGLGRH 71
DB 14 RLCAALGLTFSKILESEGLYGIC--TPNKK---VSQYRIKL-----EEROKLRFH 58
QY 72 Y--SQNDLVYSPTTEKHLIDGMVYRELCSAITHSDNINANL-LTTTGGPRELTAFLH 128
DB 59 YGITEHQQLKLYVRAR--ANGSGQILISQLEMLRDNITIFRLGKSPPTPGAGQLVNHRH 116
QY 129 NM-GDHWTRIDRWEPELNEAIPNDEPDTYMPVAMATTKLITLGLLETLIAS 178
DB 117 ILINDNTVDIPINCEPRDVIITVNNKRESVLIKNDSSKRPVPHNLTFDS 167

RESULT 11
HSLV_THEME
ID HSLV_THEME STANDARD; PRT; 176 AA.
AC Q9WY21;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR TM0521.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=MSB / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Cottonback T.R., Malek J.A., Linner K.D., Garrett M.M.,
RA Stewart A.M., Utterback M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT *Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.*
RL Nature 399:323-329(1999).
CC -1 FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1 SUBUNIT: INTERACTS WITH HSLU (By similarity).
CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).

```


DB 105 MAICECHOLVGDHDELTKNLTADVATLEYLUKVL 144

RESULT 13

YDEL_ECOLI STANDARD; PRT; 172 AA.

ID YDEL_ECOLI
AC P31131
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ydel.
CN YDEL OR B1537.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93186717; PubMed=8383113;
RA Cohen S.P., Haeckler H., Levy S.B.;
RT "Genetic and functional analysis of the multiple antibiotic
resistance (mar) locus in Escherichia coli.";
RT J. Bacteriol. 175:1484-1492(1993).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RA MEDLINE=9751357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasei H., Kashino K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RT DNA Res. 3:363-377(1996).
RL [1]
CC -1- SIMILARITY: BELONGS TO THE CINA FAMILY. STRONG, TO E.COLI YGAD.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

CC EMBL: M96235; NOT_ANNOTATED_CDS.
DR EMBL: AE000251; AAC74610.1;
DR EMBL: D90796; BAA15227.1;
DR EMBL: D90797; BAA15239.1;
DR PIR: D64908; D64908.
DR Ecogen: EG1645; ydel.
DR HAMAP: MF_00226; atypical; 1.
DR InterPro: IPR003396; CINA.
DR Pfam: PF02464; CINA; 1.
DR TIGRfam: TIGR00199; cina_cterm; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 172 AA; 18321 MW; 700A7A95C967C868 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 172;
Best Local Similarity 27.1%; Pred. No. 3.1e+02;

Matches 23; Conservative 9; Mismatches 26; Indels 27; Gaps 3;

OY 73 QNDLVESPYVEKHLTDGTVRELCSAATMSDNTANLLTIGP----- 120
DB 79 QOOSLERYSAVER-----VAAEMATGATERAD-ADVSAITGYGGEEDGTPACTV 131
OY 121 -----KELTAFLNMGDHYRL 137
DB 132 WFAMIKGONTYAMHFAQDCETVL 156

RESULT 14

PYRE_STRCO STANDARD; PRT; 182 AA.

ID PYRE_STRCO
AC Q9X8R7
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Orotate phosphoribosyltransferase (EC 2.4.2.10) (OPRT) (OPRTase).
CN PYRE OR SC03650 OR SCH10.28C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2) / M145;
RC MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Kieser H.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RT Nature 417:141-147(2002).
RL [1]
CC -1- CATALYTIC ACTIVITY: Orotidine 5'-phosphate + diphosphate -> orotate
+ 5-phospho-alpha-D-ribose 1-diphosphate.
CC -1- COFACTOR: Magnesium (by similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; fifth step.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
PHOSPHORIBOSYLTRANSFERASE FAMILY. PYRE SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)

DR EMBL: AL939117; CAB42037.1;
DR PIR: T36540; T36540.
DR HAMAP: MF_01208; 1.
DR InterPro: IPR004467; Or-phospho-trans.
DR InterPro: IPR002375; py/py-rp-transf.
DR InterPro: IPR000836; PRTtransferase.
DR Pfam: PF00156; Pribosyltran; 1.
DR TIGRfam: TIGR00336; pyre; 1.
DR PROSITE: PS00103; PUR_PYR_PR_TRANSFER; FALSE_NEG.
KW Pyrimidine biosynthesis; Transferase; Glycosyltransferase; Magnesium;
KW Complete proteome.
FT DOMAIN 118 130 PRP-BINDING (BY SIMILARITY).
FT ACT_SITE 100 100 BY SIMILARITY.
SQ SEQUENCE 182 AA; 19082 MW; FD5D4A3089137BC CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 182;
Best Local Similarity 28.8%; Pred. No. 3.3e+02;

Matches 34; Conservative 14; Mismatches 41; Indels 29; Gaps 7;

OY 165 LRRL-LTGGELTLASROOLID-----WMEDKVGPIRLSALPAG---WFI 206
 DB 36 LRRTITDGEAPLVG-QVLLDITLADLEPDANGLTGACPVAAASHMAAARRRLDAFY 94
 OY 207 ADKSGAGGERSGRIIAGDPGKPSRIYVYTT---GSOATMDEHNRQIAEGASLI 260
 DB 95 VRRTAKAHGIGRRY---EGPEIKGRRVVVEDTSTGSGPLTAVE---AVREAGAEVY 146

RESULT 15

TRPF_THEVO
 ID TRPF_THEVO STANDARD; PRT; 195 AA.
 AC 0979V6;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).
 GN TRP OR TV1054 OR TVG1080245.
 OS Thermoplasma volcanium.
 OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
 OC Thermoplasmataceae; Thermoplasmata.
 OX NCBI_TaxID=50339;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN-GSS1 / DSM 4299 / JCM 9571;
 RX MEDLINE-20570466; PubMed-11121031;
 RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
 Kawashima Y., Watanabe Y., Aramaki H., Kanehori K., Kawamoto T.,
 Nunoshiba T., Yamamoto Y., Makino K., Suzuki M.;
 "Archaeal adaptation to higher temperatures revealed by genomic
 RT sequence of Thermoplasma volcanium";
 Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
 RL -1- CATALYTIC ACTIVITY: N-(5'-phospho-beta-D-riboseyl)-anthranilate - 1-
 (2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate.
 CC -1- PATHWAY: tryptophan biosynthesis; third step.
 CC -1- SIMILARITY: BELONGS TO THE TRPF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; AP000995; BAB60196.1; -;
 DR HAMAP; MF_00135; -; 1.
 DR InterPro: IPR001240; PRAI.
 DR Pfam; PF00697; PRAI; 1.
 KM isomerase; tryptophan biosynthesis; Complete proteome.
 SO SEQUENCE 195 AA; 21325 MW; FE3C91205EE080D3 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 195;

Best Local Similarity 27.0%; Pred. No. 3.6e+02;

Matches 51; Conservative 21; Mismatches 90; Indels 27; Gaps 9;

OY 48 KVLGCAVLSRIDAGOEOLGRRIHYSONDLVEYSPVTEKHLTDGMTVRELCSAATMSDN 107
 DB 2 KIKVCGITRLEDAAVTELGAIVGVLD--ELSPRIASH-----NTIREIAEGITVAGV 55
 OY 108 TAANILLTTIGPK-ELTAFLLNMGDVTRLDRWEPELNAIPNDRDTPMPVAMATTLR 166
 DB 56 YTSE--QTVLSPLEDEYVQLH--FDH-----DPELIRSIHELGR-KVTSVINFGIR 103
 OY 167 KLUGELTLASROOLIDMEADKVGPIRLSALPAGWFIADKSGAGERSGRIIAGLP 226
 DB 104 DL---KIKYNAVREADIILVERKGVSSIVSQIAPGLANVGAGISDRDIENITIAA--- 157
 OY 227 DGKPSRIYV 235
 DB 158 --KPSIIDY 164

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_30E
Perfect score: 1348
Sequence: 1 HPETLVKVAEDQLGARVQ.....TMDERNRQIAETGASLIKHV 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_TREMBL_23:*
2: SP_Archea:*
3: SP_Fungi:*
4: SP_Human:*
5: SP_Invertebrate:*
6: SP_Mammal:*
7: SP_Mhc:*
8: SP_Organelle:*
9: SP_Phage:*
10: SP_Plant:*
11: SP_Rodent:*
12: SP_Virus:*
13: SP_Vertebrate:*
14: SP_Unclassified:*
15: SP_Virus:*
16: SP_Bacteriap:*
17: SP_Archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 38.9 | 105 | 2 | Q52026 |
| 2 | 524 | 38.9 | 105 | 2 | Q52026 |
| 3 | 493 | 36.6 | 145 | 2 | Q8RTD8 |
| 4 | 488 | 36.2 | 102 | 2 | Q52639 |
| 5 | 458 | 34.0 | 138 | 2 | Q9GQF3 |
| 6 | 458 | 34.0 | 138 | 2 | Q9AMAI |
| 7 | 458 | 34.0 | 139 | 2 | Q9AMA2 |
| 8 | 456 | 33.8 | 138 | 2 | Q9AMAO |
| 9 | 456 | 33.8 | 138 | 2 | Q9AM99 |
| 10 | 456 | 33.8 | 139 | 2 | Q9AM98 |
| 11 | 453 | 33.6 | 139 | 2 | Q9AM93 |
| 12 | 338 | 25.1 | 67 | 2 | Q53553 |
| 13 | 320 | 23.7 | 95 | 2 | Q8GDES |
| 14 | 316 | 23.4 | 62 | 2 | Q9JN58 |
| 15 | 299 | 22.2 | 128 | 2 | Q8VQ00 |
| 16 | 246 | 18.2 | 52 | 2 | Q9R412 |

| | | | | | | |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 245 | 18.2 | 180 | 2 | Q8KVT2 | Q8KVT2 staphylococ |
| 18 | 241 | 17.9 | 48 | 2 | Q9RLH0 | Q9RLH0 proteus mir |
| 19 | 197 | 14.6 | 38 | 2 | P97145 | P97145 escherichia |
| 20 | 164.5 | 12.2 | 134 | 2 | Q8VUL3 | Q8VUL3 staphylococ |
| 21 | 119.5 | 8.9 | 109 | 2 | Q53698 | Q53698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | P97146 | P97146 escherichia |
| 23 | 87 | 6.5 | 100 | 2 | Q93S05 | Q93S05 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | Q9XBJ2 | Q9XBJ2 bacillus ce |
| 25 | 79 | 5.9 | 198 | 2 | Q9ACM8 | Q9ACM8 streptococ |
| 26 | 77 | 5.7 | 181 | 16 | Q8D108 | Q8D108 yersinia pe |
| 27 | 75.5 | 5.6 | 190 | 5 | Q76227 | Q76227 trypanosoma |
| 28 | 75.5 | 5.6 | 200 | 16 | Q97JEB | Q97JEB clostridium |
| 29 | 75.5 | 5.6 | 202 | 16 | Q98520 | Q98520 rhizobium l |
| 30 | 74 | 5.5 | 177 | 16 | Q9KEM7 | Q9KEM7 bacillus ha |
| 31 | 73.5 | 5.5 | 195 | 16 | Q9RI40 | Q9RI40 streptomyce |
| 32 | 73.5 | 5.5 | 204 | 11 | Q91YV4 | Q91YV4 mus musculu |
| 33 | 73 | 5.4 | 153 | 5 | Q8S520 | Q8S520 clona lites |
| 34 | 73 | 5.4 | 175 | 17 | Q82TH8 | Q82TH8 pyrobaculum |
| 35 | 72.5 | 5.4 | 113 | 2 | Q49970 | Q49970 mycobacteri |
| 36 | 72.5 | 5.4 | 145 | 2 | Q05984 | Q05984 staphylococ |
| 37 | 71.5 | 5.3 | 152 | 16 | Q9RI91 | Q9RI91 streptomyce |
| 38 | 71 | 5.3 | 131 | 2 | Q9X9H0 | Q9X9H0 yersinia en |
| 39 | 71 | 5.3 | 169 | 17 | Q972W2 | Q972W2 sulfolobus |
| 40 | 71 | 5.3 | 192 | 16 | Q9HW07 | Q9HW07 pseudomonas |
| 41 | 71 | 5.3 | 203 | 4 | Q96B23 | Q96B23 homo sapien |
| 42 | 70.5 | 5.2 | 131 | 16 | Q8YOL8 | Q8YOL8 anabena sp |
| 43 | 70.5 | 5.2 | 184 | 5 | Q45379 | Q45379 caenorhabdl |
| 44 | 70 | 5.2 | 108 | 17 | Q9HRO3 | Q9HRO3 halobacteri |
| 45 | 70 | 5.2 | 150 | 10 | Q941G4 | Q941G4 nicotiana t |

ALIGNMENTS

RESULT 1

Q52026 PRELIMINARY; PRT; 105 AA.
ID Q52026;
AC Q52026;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN bla.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RX MEDLINE=95011664; PubMed=7926843;
RA West S.E., Schweizer H.P., Dail C., Sample A.K., Runyen-Janecky L.J.;
RT "Construction of Improved Escherichia-Pseudomonas shuttle vectors
RT derived from pUC18/19 and sequence of the region required for their
RT replication in Pseudomonas aeruginosa.";
RL Gene 148:81-86(1994).
DR EMBL: L30112; AAA65058.1; -
DR HSSP: P00810; IXPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta_lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER.
SQ SEQUENCE 105 AA; 11229 MW; D2889A4073330557 CRC64;

Query Match 38.9%; Score 524; DB 2; Length 105;
Best local similarity 100.0%; Pred. No. 1e-36;

Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AAATTAKLLTGELLTASROOLDIMWEADKVAAGPLLSALPAGWFIADKSGGERSRG 219
DB 2 AAATTLKLLTGELLTASROOLDIMWEADKVAAGPLLSALPAGWFIADKSGGERSRG 61

OY 220 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKW 263
 DB 62 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKW 105

RESULT 2

ID 052330 PRELIMINARY; PRT; 105 AA.
 AC 052330:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Tnl bla protein (Fragment).
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90264294; Pubmed=2160936;
 RA Kornacki J.A., Burlage R.S., Figurski D.H.;
 RT "The *kil*-kor regulation of broad host-range plasmid RK2: Nucleotide
 RT sequence, polypeptide product and expression of regulatory gene
 RT *korc*.";
 RL J. Bacteriol. 173:3040-3050(1990).
 DR EMBL: M32794; AA26408.1; -.
 DR HSSP: P00810; 1XPB.
 DR Interpro: IPR001466; Beta_Lactamase.
 DR Interpro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW plasmid.
 FT NON_TER 1 1
 SQ SEQUENCE 105 AA; 11229 MW; D28894A07330557 CRC64;

Query Match 38.9%; Score 524; DB 2; Length 105;
 Best Local Similarity 100.0%; Pred. No. 1e-36;
 Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 219
 DB 2 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 61
 OY 220 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKW 263
 DB 62 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKW 105

RESULT 3

ID 08RTD8 PRELIMINARY; PRT; 145 AA.
 AC 08RTD8:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE SHV-5 enzyme (Fragment).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Palaustramantam S.;
 RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
 RT ceftazidime-resistant *Klebsiella pneumoniae*.";
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF467105; AAL75506.1; -.
 DR Interpro: IPR001466; Beta_Lactamase.
 DR Pfam: PF00144; beta-lactamase; 1.
 FT NON_TER 1 1
 SQ SEQUENCE 145 AA; 145 MW; DA95AF7557DA13D4 CRC64;

OY 160 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 219
 DB 2 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 61
 OY 220 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKW 263
 DB 62 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGASLIRKW 105

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 36.6%; Score 493; DB 2; Length 145;
 Best Local Similarity 68.1%; Pred. No. 6.7e-34;
 Matches 94; Conservative 18; Mismatches 26; Indels 0; Gaps 0;

OY 100 AATMSDVTANLLTITGGPKELTFALHMGSHVTRLDWPELTAIPNDEADDTTPV 159
 DB 7 AAYMSDSNANLLATVGGPAGLTAFLRQIGDVTRELDMWETLEALPEDADDTTPA 66
 OY 160 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 219
 DB 67 SMAATTKRLITSORLSARSQROLQWMDRVAGPLIRSYLPAGWFIADKSGAGSGSRG 126
 OY 220 IIALGPDGKPSRIIVY 237
 DB 127 IVALGPNKAKRIIVY 144

RESULT 4

ID 052639 PRELIMINARY; PRT; 102 AA.
 AC 052639:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Bla-lactamase (Fragment).
 GN BLA.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-Tn1.
 RX MEDLINE=9436757; Pubmed=8058819;
 RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
 RA Kropinski A.M.;
 RT "Deletion and transposon mutagenesis and sequence analysis of the *poF*
 RT *pro1600* OriR region found in the broad-host-range plasmids of the *poF*
 RT series.";
 RL Plasmid 31:265-274(1994).
 DR EMBL: L22691; AAA98312.1; -.
 DR HSSP: P00810; 1XPB.
 DR Interpro: IPR001466; Beta_Lactamase.
 DR Interpro: IPR000871; Beta_Lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW plasmid.
 FT NON_TER 1 1
 SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA13D4 CRC64;

Query Match 36.2%; Score 488; DB 2; Length 102;
 Best Local Similarity 100.0%; Pred. No. 1.1e-33;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 219
 DB 2 AAMATTKRLITGELLTASRQQLIDWMEADKVAAGPLRSALPAGWFIADKSGAGSGSRG 61
 OY 220 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGA 257
 DB 62 IIALGPDGKPSRIIVYITGSOATMDERNROIAEIGA 99

RESULT 5

ID 0990F3 PRELIMINARY; PRT; 138 AA.

AC 0990F3:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

| | | | | | |
|------|-----------------------|--|--------------------|----------------|------------------|
| | Query Match | 34.0% | Score 458; | DB 2; | Length 138; |
| | Best Local Similarity | 67.7% | Pred. No. 5,6e-31; | | |
| | Matches | 86; | Conservative 17; | Mismatches 24; | Indels 0; Gaps 0 |
| OY . | 111 | NLLTITGGREKLETAFLNNMDHVTRLRWEPELNEAIPNDERDGTETMPVAMATLRKLTT | 170 | | |
| | | : : | | | |
| Dd | 10 | NLLIATVGSGAGLTAFRLQIDGNTATRLDRMETELNHALGODADDTTPASMAATLRKLTT | 69 | | |
| OY | 171 | GELLTLASROQLIDWMEADRKYAGPLLRSALPAGWFIADKSGAGSRGIITAAIGPOGRP | 230 | | |
| | | : : : : : : : | | | |
| Dd | 70 | SORLSARSQRQLQNMWDVRVAGPLIRSVLPAGWFIADRTGASKRGARGIVALLGPNNKA | 129 | | |
| OY | 231 | SRIVYY | 237 | | |
| | | | | | |
| Dd | 130 | ERIVYY | 136 | | |

| | | | | |
|--|---|---|---|---------|
| Oy | | 111 | NLLLTITGGKPELTAFLHNMGDHYTLDREPELNALPNDERDTPVAAATTLRLKLT | 170 |
| Dd | | 10 | NLLLATAGGPGGLAFLRQIGDNVTIRDLRWETELNEALPRADRTTPASMAATLRKLTT | 69 |
| Oy | | 171 | GELLTASROQLIDMMEADVKGPIILRSALPWGFIAIDKSAGAGEGRSGIITALGPDGKP | 230 |
| Dd | | 70 | SQRLSARSORQLDMWDDRVAGPLIRSVLPAGWFIADKTGASKRGARGIALLGPNNKA | 129 |
| Oy | | 231 | SRIVVIY 237 | |
| Dd | | 130 | ERIVVIY 136 | |
| RESULT 7 | | | | |
| ID | O9AMA2 | PRELIMINARY: | PRT: | 139 AA. |
| AC | O9AMA2: | | | |
| Dt | 01-JUN-2001 (TREMBLrel. 17, Created) | | | |
| Dt | 01-JUN-2001 (TREMBLrel. 17, Last sequence update) | | | |
| Dt | 01-MAR-2002 (TREMBLrel. 20, Last annotation update) | | | |
| DE | SHV beta-lactamase (Fragment). | | | |
| OS | Escherichia coli. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | | |
| CC | Enterobacteriaceae; Escherichia. | | | |
| OX | NCBI_Taxid=562; | | | |
| RP | [1] | | | |
| RC | SEQUENCE FROM N.A. | | | |
| RA | STRAIN=918; | | | |
| RA | Substratum G., Navarathnam P.; | | | |
| RT | *SHV gene sequences from ceftazidime-resistant Escherichia coli | | | |
| RT | isolates from UMC, Malaysia.*; | | | |
| RL | Submitted (DEC-2000) to the EMBL/Genbank/DDJ databases. | | | |
| DR | EMBL: AF327048; AK07463.1; . | | | |
| DR | HSSP: PI4557; ISHV. | | | |
| DR | InterPro: IPR001466; Beta_lactamase. | | | |
| DR | InterPro: IPR000871; Beta_lactamase_A. | | | |
| DR | Pfam: PF00144; beta-lactamase; 1. | | | |
| DR | PRINTS: PR00118; BLACTMASEA. | | | |
| FT | NON_TER | 1 | | |
| FT | NON_TER | 139 | | |
| FT | NON_TER | 1 | | |
| SEQ | SEQUENCE | 139 AA; 15234 MW; 0B161D09282442A8 CRC64; | | |
| Query Match 34.0% Score 458; DB 2; Length 139; | | | | |
| Matches 86; Conservative 17; Mismatches 24; Indels 0; Gaps 0 | | | | |
| Oy | | 111 | NLLLTITGGKPELTAFLHNMGDHYTLDREPELNALPNDERDTPVAAATTLRLKLT | 170 |
| Dd | | 10 | NLLLATAGGPGGLAFLRQIGDNVTIRDLRWETELNEALPCDARDTTPASMAATLRKLTT | 69 |
| Oy | | 171 | GELLTASROQLIDMMEADVKGPIILRSALPWGFIAIDKSAGAGEGRSGIITALGPDGKP | 230 |
| Dd | | 70 | SQRLSARSORQLDMWDDRVAGPLIRSVLPAGWFIADKTGASKRGARGIALLGPNNKA | 129 |
| Oy | | 231 | SRIVVIY 237 | |
| Dd | | 130 | ERIVVIY 136 | |
| RESULT 8 | | | | |
| ID | O9AMA0 | PRELIMINARY: | PRT: | 138 AA. |
| AC | O9AMA0: | | | |
| Dt | 01-JUN-2001 (TREMBLrel. 17, Created) | | | |
| Dt | 01-JUN-2001 (TREMBLrel. 17, Last sequence update) | | | |
| Dt | 01-MAR-2002 (TREMBLrel. 20, Last annotation update) | | | |
| DE | SHV beta-lactamase (Fragment). | | | |
| OS | Escherichia coli. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | | |
| CC | Enterobacteriaceae; Escherichia. | | | |
| OX | NCBI_Taxid=562; | | | |
| RP | [1] | | | |

RC STRAIN-E/98 9-1;
 RA Subramaniam G., Navaratnam P.;
 RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.";
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327052; AK07467.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 138
 SQ SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Query Match 33.8%; Score 456; DB 2; Length 138;
 Best Local Similarity 66.9%; Pred. No. 8.3e-31;
 Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTGGPRELTAFLHNGDHYTRLDREPELNEALPNDERDTTPVAMATTLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRGIDNVTRLDRMETELNEALPGARDTTTPASMAATLRLT 69
 OY 171 GELLITLARSQQLIDMEADKYAGPLLSALPAGWFIADKSGAGSGRSRIITAAIGPDGKP 230
 DB 70 SQRLSARSQRLQMWVDVAVAGPLIRSVLPAGWFIADKSGAGSGRGIIVALLGPNKKA 129
 OY 231 SRIVIVY 237
 DB 130 ERIVLY 136

RESULT 9

O9AM99 PRELIMINARY; PRT: 138 AA.

ID O9AM99
 AC O9AM99;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E/98 4-1;
 RA Subramaniam G., Navaratnam P.;
 RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.";
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327053; AK07468.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 138
 SQ SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 33.8%; Score 456; DB 2; Length 138;
 Best Local Similarity 66.9%; Pred. No. 8.3e-31;
 Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTGGPRELTAFLHNGDHYTRLDREPELNEALPNDERDTTPVAMATTLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRGIDNVTRLDRMETELNEALPGARDTTTPASMAATLRLT 69
 OY 171 GELLITLARSQQLIDMEADKYAGPLLSALPAGWFIADKSGAGSGRSRIITAAIGPDGKP 230
 DB 70 SQRLSARSQRLQMWVDVAVAGPLIRSVLPAGWFIADKSGAGSGRGIIVALLGPNKKA 129

OY 231 SRIVIVY 237
 DB 130 ERIVLY 136

RESULT 10

O9AM98 PRELIMINARY; PRT: 139 AA.

ID O9AM98
 AC O9AM98;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-E/98 3-2;
 RA Subramaniam G., Navaratnam P.;
 RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.";
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327054; AK07469.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 139
 SQ SEQUENCE 139 AA; 15234 MW; 03361A792F5442A8 CRC64;

Query Match 33.8%; Score 456; DB 2; Length 139;
 Best Local Similarity 66.9%; Pred. No. 8.4e-31;
 Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

OY 111 NLLTTGGPRELTAFLHNGDHYTRLDREPELNEALPNDERDTTPVAMATTLKLLT 170
 DB 10 NLLATVGGPAGLTAFLRGIDNVTRLDRMETELNEALPGARDTTTPASMAATLRLT 69
 OY 171 GELLITLARSQQLIDMEADKYAGPLLSALPAGWFIADKSGAGSGRSRIITAAIGPDGKP 230
 DB 70 SQRLSARSQRLQMWVDVAVAGPLIRSVLPAGWFIADKSGAGSGRGIIVALLGPNKKA 129
 OY 231 SRIVIVY 237
 DB 130 ERIVLY 136

RESULT 11

O9AM93 PRELIMINARY; PRT: 139 AA.

ID O9AM93
 AC O9AM93;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE SHV beta-lactamase (Fragment).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-935;
 RA Subramaniam G., Navaratnam P.;
 RT *SHV gene sequences from ceftriaxime-resistant *Escherichia coli*
 RL Isolates from UMMC, Malaysia.";
 DR Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF327047; AK07462.1; -.
 DR HSSP: P14557; ISHV.
 DR InterPro: IPR001466; Beta_lactamase.

DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 FT NON_TER 1
 FT NON_TER 139 139
 SQ SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Query Match
 Best Local Similarity 66.9%; Score 453; DB 2; Length 139;
 Matches 85; Conservative 18; Mismatches 24; Indels 0; Gaps 0;

QY 111 NLLTTTGPKKELTALHNKGHTRLDRWPEPELNEAIPNDERDTTPVAMATTLRLKLT 170
 Db 10 HLLATVGGPAGITLAFRLQIGDVTRIDRWETELNEALPGDARDTTPASMAATLRLKLT 69
 QY 171 GELLTLASRQQLIDWMEADRVAGPLRSALPAGWFIADKSGAGRGSGITIALGPDGKP 230
 Db 70 SORLSARSORQLQMWVDRVAGPLIRSVLPAGWFIADKSGARKGARGIVALLGPNRKA 129
 QY 231 SRIIVYI 237
 Db 130 ERIIVYI 136

RESULT 12

Q53553 PRELIMINARY; PRT; 67 AA.

AC 053553:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE TEM-type beta-lactamase (Fragment).
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UCSM 129;
 RX MEDLINE=96081517; PubMed=8538491;
 RA Echeverria V., Olate J., Cid H.;
 RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
 flexneri strain."
 RL Microbios 83:107-117(1995).
 DR EMBL: S81098; AAB35839.2; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR Pfam: PF00144; beta-lactamase; 1.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER 1
 FT NON_TER 67
 SQ SEQUENCE 67 AA; 7287 MW; 85F8C85B0363FCB CRC64;

Query Match
 Best Local Similarity 25.1%; Score 338; DB 2; Length 67;
 Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 55 VLSRIIDGQEOELGRIRHSNDLVESPTKRLTNGMTVRELCSAAITMSDNTAANLL 114
 Db 1 VLSHVDAGQEOELGRIRHSNDLVESPTKRLTNGMTVRELCSAAITMSDNTAANLL 60
 QY 115 TTIGGPK 121
 Db 61 TTIGGPK 67

RESULT 13
 Q8GDES PRELIMINARY; PRT; 95 AA.
 AC 08GDES:

DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Extended-spectrum beta-lactamase SHV-39 (Fragment).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bello H.M., Dominguez M.P., Daashtl A.A., Gonzalez-Rocha G.E.,
 RA Ameyes S.G.B.;
 RT "SHV-39: a new extended-spectrum beta-lactamase found throughout
 Chile."
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY150585; AAN77730.1; -.
 FT NON_TER 1
 FT NON_TER 95
 SQ SEQUENCE 95 AA; 10299 MW; 496837847670413C CRC64;

Query Match
 Best Local Similarity 23.7%; Score 320; DB 2; Length 95;
 Matches 61; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

QY 147 AIPNDERDTTPVAMATTLRLKLTGELLTLASRQQLIDWMEADRVAGPLRSALPAGWFI 206
 Db 1 ALPGDARDTTPASMAATLRLKLTSSORLSARSORQLQMWVDRVAGPLIRSVLPAGWFI 60
 QY 207 ADKRGAGRGSRGITIALGPDGKPSRIVYITGGS 241
 Db 61 ADKTAGSARGARGIVALLGPNKARIVYIYIDS 95

RESULT 14

Q9JWS8 PRELIMINARY; PRT; 62 AA.

AC Q9JWS8:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE TEM-type beta-lactamase (Fragment).
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Shigella.
 OX NCBI_TaxID=623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UCSM 129;
 RX MEDLINE=96081517; PubMed=8538491;
 RA Echeverria V., Olate J., Cid H.;
 RT "Partial DNA sequence of a beta-lactamase produced by a Shigella
 flexneri strain."
 RL Microbios 83:107-117(1995).
 DR EMBL: S81099; AAB35840.1; -.
 DR HSSP: P00810; 1XPB.
 DR InterPro: IPR000871; Beta_lactamase_A.
 DR PRINTS: PR00118; BLACTAMASEA.
 KW Plasmid.
 FT NON_TER 1
 FT NON_TER 62
 SQ SEQUENCE 62 AA; 7049 MW; 1806CA19D241540E CRC64;

Query Match
 Best Local Similarity 23.4%; Score 316; DB 2; Length 62;
 Matches 61; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 132 DHTYRLDRWPEPELNEAIPNDERDTTPVAMATTLRLKLTGELLTLASRQQLIDWMEADRV 191
 Db 1 DHTYRLDRWPEPELNEAIPNDERDTTPVAMATTLRLKLTGELLTLASRQQLIDWMEADRV 60
 QY 192 AG 193
 ||

Db 61 AG 62

RESULT 15

ID Q8VQ00 PRELIMINARY; PRT; 128 AA.

AC QBVQ00
DT 01-MAR

| | |
|----|--|
| DI | 01-MAR-2002 (11 Emblicel: 20, created) |
| DT | 01-MAR-2002 (Tremblay: 20, last seq) |

DT 01-MAR-2003 (TREMBLÉ, 23, last annotation update)

CTX-M type beta-lactamase (Fragment).

05 *Escherichia coli*.
00 *Bacteriophage: Gamma*

Enterobacteriaceae; Escherichia.

NCBI_TaxID=562;

RN [2]

| RP | SEQ |
|----|-----|
| PC | STI |

RA Stu

RT "A new CTX-M type beta-lactamase

Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DB InterPro: TPB001466: Beta lactamase;
DR EMBL; AF406148; AACL/3903.1; -;

DR InterPro; IPR000871; Beta_lactamase

DR Pfam; PF00144; beta-lactamase; 1.

DR PRINTS; PROUIB; BLACIAMASEA.
ET NON TER 1

| ET | NON-TER | 1 | 128 |
|----|---------|---|-----|
| ET | NON-TER | 1 | 128 |

| | | | | | | |
|----|----------|-----|-----|-------|-----|--------|
| 5Q | SEQUENCE | 128 | AA; | 13878 | MM; | F2EBB6 |
|----|----------|-----|-----|-------|-----|--------|

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

| | |
|-----------------------|--------|
| Query Match | 22.26% |
| Best Local Similarity | 48.48% |

Best book primarily
Matches 59; Conservat

| Query Match | 22.28; | Score 299; | DB 2; | Length 128; |
|-------------|--------|------------|-------|-------------|
|-------------|--------|------------|-------|-------------|

Best Local Similarity 48.48; Pred. No. 1.3e-17;
Matches 59; Conservative 19; Mismatches 44; Indels 0; Gaps 0;

[illegible]

66 LGRIHYSQNDLVEYSPVTEKHLTDGMTVRELCSAITMSDNTAANLLTTIGPKELTA 125

[illegible]

DD 0 LNDKVEIKNSDLVNINP IAEKNAVNG IWSLWELSDZKVALQ I SDNVKHEUNNKL I KHVVOGFADVA IA 00

126 FLHMGDHTRLDRWEPLENAIPNDERDTMPVAMATTLRKLITGELLTIASRQQLIDW 185

[illegible]

DB 66 FARU LGDEI FR LDR I E M L N I A L P G D P R D I T S P R A M A Q I L R N L I U S K A L G D S Q K A Q L V I W I Z

Search completed: September 10, 2003, 12:29:26
Job time : 26.4286 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 32.7143 Seconds
(without alignments)
1276.051 Million cell updates/sec

Title: SEQ2_30E_157T
1348

Sequence: 1 HPELVKVKADIEDQLGARVG.....TMDERNQIAEIGASLIKHW 263

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 865366

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_19Jun03:*

| | |
|-----|--|
| 1: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:* |
| 2: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:* |
| 3: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:* |
| 4: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:* |
| 5: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:* |
| 6: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:* |
| 7: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:* |
| 8: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:* |
| 9: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:* |
| 10: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:* |
| 11: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:* |
| 12: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:* |
| 13: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:* |
| 14: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:* |
| 15: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:* |
| 16: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:* |
| 17: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:* |
| 18: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:* |
| 19: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:* |
| 20: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:* |
| 21: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:* |
| 22: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:* |
| 23: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:* |
| 24: | /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 587 | 43.5 | 196 | 7 AAP60627 | Beta-urogastrone - |
| 2 | 571 | 42.4 | 146 | 22 AAU23221 | Novel human enzyme |
| 3 | 571 | 42.4 | 146 | 22 AAM90870 | Human immune enzyme |
| 4 | 492 | 36.5 | 127 | 18 AAU20440 | H. pylori cytolins |
| 5 | 477 | 35.4 | 129 | 22 AAU23220 | Novel human enzyme |
| 6 | 477 | 35.4 | 129 | 22 AAM90871 | Human immune/haema |
| 7 | 477 | 35.4 | 182 | 22 ABG27917 | Novel human diagno |
| 8 | 457 | 33.9 | 94 | 21 AAB59052 | Breast and ovarian |
| 9 | 404 | 30.0 | 159 | 7 AAP60628 | Beta-urogastrone - |

| | | | | | |
|----|------|------|-----|----|-----------|
| 10 | 366 | 27.2 | 101 | 22 | ABG27935 |
| 11 | 333 | 24.7 | 88 | 22 | ABG27919 |
| 12 | 199 | 14.8 | 119 | 7 | AAP60626 |
| 13 | 141 | 10.5 | 51 | 21 | AAU2783 |
| 14 | 80.5 | 6.0 | 158 | 22 | AAU45138 |
| 15 | 78 | 5.8 | 15 | 10 | AAU98503 |
| 16 | 77 | 5.7 | 200 | 18 | AAU55524 |
| 17 | 75.5 | 5.6 | 170 | 19 | AAU38722 |
| 18 | 75 | 5.6 | 202 | 24 | AAU16283 |
| 19 | 75 | 5.6 | 202 | 24 | AAU16283 |
| 20 | 71.5 | 5.3 | 196 | 21 | AAU73344 |
| 21 | 71 | 5.3 | 159 | 19 | AAU69170 |
| 22 | 70.5 | 5.2 | 193 | 23 | AAU76854 |
| 23 | 70.5 | 5.2 | 193 | 23 | AAU76853 |
| 24 | 70 | 5.2 | 170 | 22 | AAU52472 |
| 25 | 70 | 5.2 | 202 | 22 | AAU52472 |
| 26 | 69.5 | 5.2 | 170 | 22 | AAU56330 |
| 27 | 69.5 | 5.2 | 184 | 21 | AAU25446 |
| 28 | 69 | 5.1 | 134 | 21 | AAU27220 |
| 29 | 69 | 5.1 | 166 | 23 | AAU26537 |
| 30 | 69 | 5.1 | 194 | 22 | AAU81764 |
| 31 | 68 | 5.0 | 180 | 21 | AAU20206 |
| 32 | 67.5 | 5.0 | 152 | 22 | AAU19138 |
| 33 | 67.5 | 5.0 | 152 | 22 | ABG27771 |
| 34 | 67.5 | 5.0 | 187 | 21 | AAU605495 |
| 35 | 67.5 | 5.0 | 187 | 21 | AAU646629 |
| 36 | 67.5 | 5.0 | 187 | 23 | AAU97104 |
| 37 | 67 | 5.0 | 178 | 22 | AAU91926 |
| 38 | 67 | 5.0 | 196 | 18 | AAU14564 |
| 39 | 66.5 | 4.9 | 83 | 22 | AAU59092 |
| 40 | 66 | 4.9 | 183 | 19 | AAU85905 |
| 41 | 66 | 4.9 | 202 | 24 | AAU32764 |
| 42 | 65.5 | 4.9 | 50 | 17 | AAU03591 |
| 43 | 65.5 | 4.9 | 102 | 23 | AAU4479 |
| 44 | 65.5 | 4.9 | 136 | 21 | AAU44229 |
| 45 | 65.5 | 4.9 | 177 | 21 | AAU41582 |

ALIGNMENTS

| | | | | | |
|----------|---------------------------------|--|--|--|--|
| RESULT 1 | | | | | |
| ID | AAU60627 | standard; Protein; 196 AA. | | | |
| XX | AAU60627 | | | | |
| AC | AAU60627 | | | | |
| XX | 25-MAR-2003 | (updated) | | | |
| DT | 17-JUN-1991 | (first entry) | | | |
| XX | Beta-urogastrone | - beta-lactamase fusion protein from pUG2101. | | | |
| DE | Beta-urogastrone | - beta-lactamase fusion protein from pUG2101. | | | |
| XX | Beta-urogastrone | gene; hormone; salivary gland; ulcers; wounds; | | | |
| KW | fusion protein; beta-lactamase. | | | | |
| XX | Synthetic. | | | | |
| OS | Synthetic. | | | | |
| XX | Key | Location/Qualifiers | | | |
| FT | Protein | 1..120 | | | |
| FT | Misc-difference | 121..123 | | | |
| FT | Protein | /label= adaptor | | | |
| FT | Protein | 124..196 | | | |
| XX | DE3523634-A. | /label= beta-urogastrone | | | |
| PN | DE3523634-A. | | | | |
| XX | 09-JAN-1986. | | | | |
| PD | 09-JAN-1986. | | | | |
| XX | 02-JUL-1985; | 85DE-3523634. | | | |
| PF | 02-JUL-1985; | | | | |
| XX | 02-JUL-1984; | 84JP-0137691. | | | |
| PR | 02-JUL-1984; | | | | |
| XX | | | | | |

Novel human diagno
Novel human diagno
Beta-urogastrone -
Vggs-beta-lactama
Protonlactacium
Sequence encoded b
H. pylori ORF 29p
S. pneumoniae SGR
Human zcyto24 prot
Mouse IXY129840-2
HTRM clone 0258181
N-terminally tagge
Human integrin alp
Human integrin alp
Protonlactacium
S. epidermidis ope
Protonlactacium
Pinus radiata cell
Zea mays protein f
S. epidermidis ope
Arabidopsis thalia
Novel human diagno
Arabidopsis thalia
Arabidopsis thalia
Human M61 protein
C glutamicum prote
Streptococcus pneu
Protonlactacium
S. pneumoniae derl
Human zcyto25 prot
Human alpha 2 C4 a
Human dehydrotogena
Partial corn extra
Human ORFX ORF1346

20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244517.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249285.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251997.
PR 05-JAN-2001; 2000US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465566/50.
XX N-PSDB; AAS41091.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
XX preventing, treating neural, immune system, muscular, reproductive,
XX pulmonary, cardiovascular, renal, proliferative disorders and cancerous
XX diseases -
XX
XX Claim 11; SEQ ID NO 1217; 1180pp; English.

| | |
|--|--|
| CC | diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. |
| CC | AU22915-AU23814 represent the novel human enzyme polypeptides of the invention. |
| CC | Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from MRO at ftp.wipo.int/pub/published_pct_sequences. |
| CC | |
| CC | |
| XX | Sequence 146 AA; |
| SQ | |
| Query Match | 42.4%; Score 571; DB 22; Length 146; |
| Best Local Similarity | 97.4%; Pred. No. 1.5e-51; |
| Matches 112; Conservative 1; Mismatches 2; Indels 0; Gaps 0; | |
| DY | 1 HPETLVKVKDAEDQLGARVGYIELDLNSGELLESPRERFPMKSTFKVLICGVLSRID 60 DB 28 HPEITLVKVDKDIEDQLGARVGTYIEDLNISGLESPRERFPMKSTFKVLICGVLSRID 87 |
| OY | 61 AGCEOLGRRIHYSONDLVEYSPTVEKHLLTDGTVRELCSAATWSDMTAAANLLLT 115 DB 88 AGCEOLGRRIHYSONDLVEYSPTVEKHLLTDGTVRELCXAATWSDMTAVNLLLT 142 |
| RESULT 3 | |
| AAM90870 | AAM90870 standard; Protein: 146 AA. |
| ID | |
| XX | AAM90870; |
| AC | |
| XX | 07-NOV-2001 (first entry) |
| DT | |
| XX | |
| DE | Human Immune/haematopoietic antigen SEQ ID NO:18463. |
| XX | |
| KM | Human; immune; haematopoietic; immune/haematopoietic antigen; cancer; cytosstatic; gene therapy; vaccine; metastasis. |
| XX | Homo sapiens. |
| OS | |
| XX | WO200157182-A2. |
| PN | |
| PD | 09-AUG-2001. |
| XX | |
| PF | 17-JAN-2001; 2001MO-USO1354. |
| XX | |
| PR | 31-JAN-2000; 2000US-0179065. |
| PR | 04-FEB-2000; 2000US-0180268. |
| PR | 24-FEB-2000; 2000US-0184664. |
| PR | 02-MAR-2000; 2000US-0186350. |
| PR | 15-MAR-2000; 2000US-0189874. |
| PR | 17-MAR-2000; 2000US-0190076. |
| PR | 18-APR-2000; 2000US-0198123. |
| PR | 19-MAY-2000; 2000US-0205515. |
| PR | 07-JUN-2000; 2000US-0209467. |
| PR | 28-JUN-2000; 2000US-0214886. |
| PR | 30-JUN-2000; 2000US-0215135. |
| PR | 07-JUL-2000; 2000US-0216647. |
| PR | 07-JUL-2000; 2000US-0216880. |
| PR | 11-JUL-2000; 2000US-0217487. |
| PR | 11-JUL-2000; 2000US-0217496. |
| PR | 14-JUL-2000; 2000US-0218290. |
| PR | 26-JUL-2000; 2000US-0220963. |
| PR | 26-JUL-2000; 2000US-0220964. |
| PR | 14-AUG-2000; 2000US-0224518. |
| PR | 14-AUG-2000; 2000US-0224519. |
| PR | 14-AUG-2000; 2000US-0225213. |
| PR | 14-AUG-2000; 2000US-0225214. |

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226588.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236357.
PR 29-SEP-2000; 2000US-0236358.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
XX

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0251997.
PR 05-JAN-2001; 2001US-0259578.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI. 2001-483426/52.
DR N-PSDB; AAK63651.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Claim 11; SEQ ID NO 18463; 3071pp + sequence listing; English.
XX
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cystostic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX

OY 1 HPEFLVKVDAEDQLGARVGYIELDNSGELLSEFRPERPPMSTFKVLGAVLSRID 60
Db 28 HPEFLVKVDAEDQLGARVGYIELDNSGKLTLESPRPERPPMSTFKVLGAVLSRID 87
OY 61 AGQEQLGRRIRHYSQNDLVERSPYTEKHLTDGMVRE 96
Db 88 AGQEQLGRRIRHYSQNDLVERSPYTEKHLTDGMVRE 123

RESULT 6
AAM90871
ID AAM90871 standard; Protein; 129 AA.
XX
AC AAM90871;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:18464.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN MO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0239678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK63652.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
PS Claim 11; SEQ ID NO 18464; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 129 AA;
Query Match 35.4%; Score 477; DB 22; Length 129;
Best local Similarity 96.9%; Pred. No. 8.8e-42;
Matches 93; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 HPETLVKKAEDQLGARVGIETDINSGLIESFRPERFPMSTFKYLLCGAVLSRID 60
DB 28 HPETLVKKAEDQLGARVGIETDINSGLIESFRPERFPMSTFKYLLCGAVLSRID 87
OY 61 AGQEQIGRRIRHYSNDLVEYSPVTEKHLTDGMTVRE 96
DB 88 AGQEQIGRRIRHYSNDLVEYSPVTEKHLTDGMTVRE 123

RESULT 7
ABG27917
ID ABG27917 standard; Protein; 182 AA.
XX
AC ABG27917;

XX
XX 18-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #27908.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001MO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR
XX
XX (HYSE-) HYSEO INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS92104.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 20; SEQ ID NO 58276; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 182 AA;
Query Match 35.4%; Score 477; DB 22; Length 182;
Best local Similarity 87.3%; Pred. No. 1.5e-41;
Matches 96; Conservative 4; Mismatches 6; Indels 4; Gaps 1;

OY 1 HPETLVKKAEDQLGARVGIETDINSGLIESFRPERFPMSTFKYLLCGAVLSRID 60
DB 24 HPETLVKKAEDQLGARVGIETDINSGLIESFRPERFPMSTFKYLLCGAVLSRID 83
OY 61 AGQEQIGRRIRHYSNDLVEYSPVTEKHLTDGMTVRELSAATMNDNTAA 110
DB 84 AGQEQIGRRIRHYSNDLVEYSPVTEKHLTDGKTER-----LEKSFSDNTAS 129

RESULT 8
AAB59052
ID AAB59052 standard; Protein; 94 AA.

```

XX AA59052;
AC
XX
XX 27-MAR-2001 (first entry)
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 760.
XX
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.
XX
OS Homo sapiens.
XX
XX WO2005173-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05681.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-611515/58.
XX
XX N-PSDB; AAF21955.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Claim 11; Page 1228; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiant activity. The
XX polynucleotide and protein sequences are used in the diagnosis of cancer,
XX particularly breast and ovarian cancer. The nucleic acid sequences,
XX proteins, agonists and antagonists may also be used in the diagnosis,
XX prevention and treatment of immune disorders e.g. Addison's disease,
XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
XX arthritis and ulcerative colitis; cardiovascular disorders such as
XX myocardial ischaemia; wound healing; neurological diseases such as
XX cerebral anoxia and epilepsy; and infectious diseases.
XX
XX
XX Sequence 94 AA:
XX
XX Query Match 33.9%; Score 457; DB 21; Length 94;
XX Best Local Similarity 97.8%; Pred. No. 6.9e-40;
XX Matches 86; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 75 NDVEYSPYTERKHLTDGVTRELCASATMSDNTANLLTTGGRELAFLHMGDEV 134
XX |||||
XX 1 NDVEYSPYTERKHLTDGVTRELCASATMSDNTANLLTTGGRELAFLHMGDEV 60
XX |||||
XX 135 TRLDWEPELNEALPNDERDFTTPVAMATT 164
XX |||||
XX 61 TRLDWEPELNEALPNDERDFTTPVAMATT 90
XX |||||

```

```

RESULT 9
ID AAP60628 standard; Protein; 159 AA.
XX
XX AAP60628;
AC
XX
XX 25-MAR-2003 (updated)
XX 17-JUN-1991 (first entry)
XX
XX Beta-urogastrone - beta-lactamase fusion protein from pUG2701.
XX
XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
XX fusion protein; beta-lactamase.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Protein 1..96
XX /label= beta-lactamase
XX Misc-difference 97..100
XX FT /label= adaptor
XX FT 101..159
XX FT Protein /label= beta-urogastrone
XX
XX DE3523634-A.
XX
XX 09-JAN-1986.
XX
XX 02-JUL-1985; 85DE-3523634.
XX
XX 02-JUL-1984; 84JP-0137691.
XX
XX (EART) EARTH CHEM CO LTD.
XX
XX Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A,
XX Matsushiro S;
XX
XX WPI: 1986-015031/03.
XX
XX N-PSDB; AAN60632.
XX
XX New gene for expression of beta-urogastrone - its derive., plasmid(s)
XX and transformed cells contg. it.
XX
XX Disclosure; Page 59-61; 92pp; German.
XX
XX The fusion protein is less easily degraded by proteases and so
XX protects beta-urogastrone and beta-lactamase collects in the periplasm
XX of E.coli. It is therefore easy to collect and purify the product.
XX Beta-urogastrone is the hormone of the salivary glands which suppresses
XX stomach acid secretion and promotes cell growth, so is useful for
XX treating ulcers and wounds. Previously the product was obtd. only
XX in small amts. from human urine.
XX See also AAN60628, and 30-32.
XX (Updated on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 159 AA:
XX
XX Query Match 30.0%; Score 404; DB 7; Length 159;
XX Best Local Similarity 68.3%; Pred. No. 5.3e-34;
XX Matches 86; Conservative 9; Mismatches 19; Indels 12; Gaps 2;
XX
XX 1 HPETLVKKAEDQLGARVGIETLDNSGEILESRPEREPFPMSTFKYLLCAVLSRID 60
XX |||||
XX 24 HPETLVKKAEDQLGARVGIETLDNSGEILESRPEREPFPMSTFKYLLCAVLSRID 83
XX |||||
XX 61 AGGEOLGRRIHYSNDLVERS-----PTERKHLTDGVTRELCASATMSDNTANL 112
XX |||||
XX 84 AGGEOLGRRIHYSNDLVERSAKRNDSDECPISH---DGYCLHDGVCMTALDKYACNC 139
XX |||||
XX 113 LRTTIG 118
XX ::||

```

DB 140 VVGTC 145

RESULT 10
ABG27935
ID ABG27935 standard; Protein: 101 AA.
XX
AC ABG27935;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27926.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR MPI: 2001-639362/73.
DR N-PSDB; AAS92122.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 58294; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 101 AA;

Query Match 27.2%; Score 366; DB 22; Length 101;
Best Local Similarity 98.6%; Pred. No. 2.6e-30;
Matches 73; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 154 DTTTPVAMATTLRKLITGELLTLASRQOLIDWMDKAVGPIRLSALPACWFIADKSGAG 213
DB 10 DTTTPVAMATTLRKLITGELLTLASRQOLIDWMDKAVGPIRLSALPACWFIADKSGAG 69

OY 214 ERSGRITIALGPD 227
DB 70 ERSGRITIALGPD 83

RESULT 11
ABG27919
ID ABG27919 standard; Protein: 88 AA.
XX
AC ABG27919;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27910.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR MPI: 2001-639362/73.
DR N-PSDB; AAS92106.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 58278; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 88 AA;

Query Match 24.7%; Score 333; DB 22; Length 88;
Best Local Similarity 81.2%; Pred. No. 6.1e-27;
Matches 69; Conservative 3; Mismatches 3; Indels 10; Gaps 1;

OY 4 TLVKKVDAEDQIGARVGYIELDNSGELLESFRRPERRPMSTFRVLLCGAVLSRIDAQ 63
DB TLVKKVDAEDQIGARVGYIELDNSGELLESFRRPERRPMSTFRVLLCGAVLSRIDAQ 63

DB 1 TLTKVKDAEDQDGLGARVGTYIELDLNSGKLTIESFRPEFRPMSTFRKVLGAVLSRIDAGQ 60
 OY 64 EQLGRRI-----HYSONDLV 78
 ||||| | | | | |
 DB 61 EQLGRRTITLRTMYLSTHOSQSKSL 85

RESULT 12

AAP60626
 ID AAP60626 standard; Protein: 119 AA.

XX AAP60626;

XX 25-MAR-2003 (updated)

DT 17-JUN-1991 (first entry)

XX Beta-urogastrone - beta-lactamase fusion protein from pUG2301.

XX Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;

KM fusion protein; beta-lactamase.

XX Synthetic.

OS Synthetic.

XX Key

FT Protein 1.63 location/Qualifiers

FT Misc-difference 64..66 /label- beta-lactamase

FT Protein 67..119 /label- adaptor

FT Protein 67..119 /label- beta-urogastrone

PN DE3523634-A.

XX 09-JAN-1986.

PD 02-JUL-1985; 85DE-3523634.

XX 02-JUL-1984; 84JP-0137691.

XX (EART) EARTH CHEM CO LTD.

PA Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;

PI Matsushiro S;

XX WPI: 1986-015031/03.

DR N-PSDB; AAN60630.

XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)

PT and transformed cells contg. it.

XX Disclosure; Page 55; 92pp; German.

XX The fusion protein is less easily degraded by proteases and so

CC protects beta-urogastrone and beta-lactamase collects in the periplasm

CC of E.coli. It is therefore easy to collect and purify the product.

CC Beta-urogastrone is the hormone of the salivary glands which suppresses

CC stomach acid secretion and promotes cell growth, so is useful for

CC treating ulcers and wounds. Previously the product was obtd. only

CC in small amts. from human urine.

CC See also AAN60628, and 30-32.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 119 AA;

Query Match 14.8%; Score 199; DB 7; Length 119;

Best Local Similarity 97.5%; Pred. No. 1e-12;

Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HPETLVVKDAEDQDGLGARVGTYIELDLNSGKLTIESFRPEER 40

DB 24 HPETLVVKDAEDQDGLGARVGTYIELDLNSGKLTIESFRPEER 63

RESULT 13
 AAY92783
 ID AAY92783 standard; Protein: 51 AA.

XX AAY92783;

XX 29-AUG-2000 (first entry)

DT Vtgs-beta-lactamase fusion protein (partial).

XX Vtgs; vitellogenin; secretory signal sequence; gene expression;

KM oestrogen receptor binding protein; systemic circulation;

XX beta-lactamase.

XX Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX Key

FT Peptide 1..21 location/Qualifiers

FT Cleavage-site 15..16 /label- Vtgs

FT Protein 22..51 /note= "beta-lactamase mature protein"

PN W0200026366-A1.

XX 11-MAY-2000.

PD 29-OCT-1999; 99WO-SG00108.

XX 30-OCT-1998; 98US-0106426.

XX 26-OCT-1999; 99US-0426776.

XX (UYSI-) UNIV SINGAPORE NAT.

PA (LAMT/) LAM T J.

XX Ding JL, Tan NS, Ho B;

PI WPI: 2000-365615/31.

DR N-PSDB; AAA28507.

XX Isolated nucleic acid for assaying for heterologous gene expression,

PT detecting presence of compound that binds to estrogen receptor or

PT producing desired protein from host cell comprises nucleotide sequence

XX encoding secretory signal sequence

XX Example 6; Fig 14A; 73pp; English.

XX A reporter beta-lactamase system that uses the Oreochromis aureus,

CC vitellogenin secretory sequence (Vtgs), designated pBADVtgbLactKana was

CC constructed. Vtgs and variants that comprise conservative

CC replacements that retain the biological activities of directing secretion

CC of a fusion protein from a cell and cleavage of the secretory signal

CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be

CC fused to either a reporter protein or a lipopolysaccharide-binding

CC protein coding sequence. The isolated nucleic acid is useful for assaying

CC for heterologous gene expression, detecting the presence of a compound

CC that binds to an oestrogen receptor in a sample or producing a desired

CC protein from a host cell. It can also be used in a method for obtaining

CC host organism.

XX Sequence 51 AA;

Query Match 10.5%; Score 141; DB 21; Length 51;

Best Local Similarity 96.6%; Pred. No. 3.6e-07;

Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 PETLVKVKDAEDQDGLGARVGTYIELDLNSGK 30

DB 23 PETLVKVKDAEDQDGLGARVGTYIELDLNSGK 51

```

RESULT 14
AAU45138
ID AAV45138 standard; protein; 158 AA.
XX
AC AAV45138;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #6034.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
  uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
  inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
  dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
  02-JUN-2000; 2000US-208841P.
  07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Metcham JL, Wang SS, Bhatia A;
  L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
  N-PSDB: AAS59525.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
  vaccinating against and diagnosing infections, especially useful for
  treating acne vulgaris.
XX
PS Example 1: SEQ ID NO 6333; 1069pp; English.
XX
CC Sequences AAU93105-AAU68017 represent Propionibacterium acnes immunogenic
  polypeptides. The proteins and their associated DNA sequences are used in
  the treatment, prevention and diagnosis of medical conditions caused by
  P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
  pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
  P. acnes is also involved in infections of bone, joints and the central
  nervous system, however it is particularly involved in the inflammatory
  lesions associated with acne vulgaris. A method for detecting the
  presence or absence of P. acnes in a patient comprises contacting a
  sample with a binding agent that binds to the proteins of the invention
  and determining the amount of bound protein in the sample. The
  polypeptides may be used as antigens in the production of antibodies
  specific for P. acnes proteins. These antibodies can be used to
  downregulate expression and activity of P. acnes polypeptides and
  therefore treat P. acnes infections. The antibodies may also be used as
  diagnostic agents for determining P. acnes presence, for example, by
  enzyme linked immunosorbent assay (ELISA).
  Note: The sequence data for this patent did not form part of the printed
  specification, but was obtained in electronic format directly from WIPO
  at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 158 AA;

Query Match 6.0%; Score 80.5; DB 22; Length 158;
Best Local Similarity 33.3%; Pred. No. 4;
Matches 26; Conservative 9; Mismatches 36; Indels 7; Gaps 2;

OY 110 ANLLTITGGREELTAFLLNMGDVTRLDWEPELNAIPNDERDITTPVAMATTLR--K 167
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80 ARMATSTSGAPRVSSASTARSGASPTVGWEPSTBPAT-----TTSAPLAFRDTLRSGR 134
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 168 LVTGELVTLASRQQLIDW 185

```

```

DB 135 GRMGSVVLATRIPIRLRW 152

RESULT 15
AAP98503
ID AAP98503 standard; protein; 15 AA.
XX
AC AAP98503;
XX
DT 26-MAR-1992 (first entry)
XX
DE Sequence encoded by probe TEM-1 which corresp. to AAs 59-73 of the
  beta-lactamase structural gene.
XX
KW Beta-lactamase; probe; antibiotic resistance; marker.
XX
OS Escherichia coli.
XX
PN EP300923-A.
XX
PD 25-JAN-1989.
XX
PF 22-JUL-1988; 88EP-0401923.
  24-JUL-1987; 87US-0077393.
XX
PR (OTLA-) UNIV LAVAL.
XX
PA Levesque R;
XX
PI WPI: 1989-026067/04.
  N-PSDB: AAN91169.
XX
DR Assaying presence of gene coding for synthesis of beta-lactamase
  by hybridisation of fixed single strand and fragments with
  labelled probe having sequence complementary to structural gene
XX
PS Disclosure: Fig 4; 21pp; English.
XX
CC The inventors claim an assay for the presence of a gene coding for
  the synthesis of beta-lactamase (BL) in a microorganism. The assay
  involves contacting a fixed single stranded DNA fragment with a
  labelled probe having a nucleotide sequence complementary to at
  least 12 bases of the BL gene.
XX
SO Sequence 15 AA;

Query Match 5.8%; Score 78; DB 10; Length 15;
Best Local Similarity 100.0%; Pred. No. 0 24;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 RPERRPPMSTFKVL 50
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 RPERRPPMSTFKVL 15

Search completed: September 10, 2003, 12:24:53
Job time : 33.7143 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 11.8571 seconds

(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_30E_157T

Perfect score: 1348

Sequence: 1 HPEPLVKKVDAEDQLGARV.....TMDERNQIAETGASLTKHW 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0

Maximum DB seq length: 207

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile91.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result | No. | Score | Query Match | Length | DB ID | Description |
|--------|------|-------|-------------|--------|----------------------|-------------------|
| 1 | 75.5 | 5.6 | 170 | 4 | US-08-858-207A-519 | Sequence 519, App |
| 2 | 74.5 | 5.5 | 124 | 4 | US-08-311-731A-202 | Sequence 202, App |
| 3 | 72.5 | 5.4 | 197 | 4 | US-09-252-991A-24241 | Sequence 24241, A |
| 4 | 71 | 5.3 | 159 | 3 | US-08-991-890-4 | Sequence 4, Appl |
| 5 | 69 | 5.1 | 169 | 4 | US-09-328-352-4491 | Sequence 4491, Ap |
| 6 | 65.5 | 4.9 | 158 | 3 | US-09-010-809-19 | Sequence 19, Appl |
| 7 | 65 | 4.8 | 150 | 3 | US-09-239-909-2 | Sequence 15, App |
| 8 | 65 | 4.8 | 181 | 3 | US-08-482-142-195 | Sequence 155, App |
| 9 | 65 | 4.8 | 181 | 3 | US-08-478-572-195 | Sequence 155, App |
| 10 | 65 | 4.8 | 181 | 3 | US-08-484-296-195 | Sequence 155, App |
| 11 | 64.5 | 4.8 | 178 | 4 | US-09-252-991A-29942 | Sequence 29942, A |
| 12 | 64 | 4.7 | 203 | 4 | US-08-624-677A-2 | Sequence 2, Appl |
| 13 | 63.5 | 4.7 | 144 | 4 | US-08-225-480-4 | Sequence 4, Appl |
| 14 | 63.5 | 4.7 | 144 | 4 | US-09-118-445-4 | Sequence 4, Appl |
| 15 | 63 | 4.7 | 174 | 2 | US-08-557-122A-6 | Sequence 6, Appl |
| 16 | 63 | 4.7 | 174 | 4 | US-09-262-666-6 | Sequence 6, Appl |
| 17 | 63 | 4.7 | 200 | 2 | US-08-557-122A-12 | Sequence 12, Appl |
| 18 | 63 | 4.7 | 200 | 4 | US-09-262-666-12 | Sequence 12, Appl |
| 19 | 62 | 4.6 | 132 | 4 | US-09-252-991A-22681 | Sequence 22681, A |
| 20 | 61.5 | 4.6 | 142 | 4 | US-09-345-473E-31 | Sequence 31, Appl |
| 21 | 61 | 4.5 | 189 | 4 | US-08-671-548C-48 | Sequence 48, Appl |
| 22 | 61 | 4.5 | 204 | 4 | US-09-252-991A-17611 | Sequence 17611, A |
| 23 | 60.5 | 4.5 | 103 | 4 | US-09-732-210-1282 | Sequence 1282, Ap |
| 24 | 60.5 | 4.5 | 167 | 2 | US-08-690-849-2 | Sequence 2, Appl |
| 25 | 60.5 | 4.5 | 167 | 3 | US-09-004-053-2 | Sequence 2, Appl |
| 26 | 60 | 4.5 | 158 | 4 | US-09-107-532A-4218 | Sequence 4218, Ap |
| 27 | 59.5 | 4.4 | 146 | 4 | US-09-134-001C-5269 | Sequence 5269, Ap |

ALIGNMENTS

| | | | | | | |
|----|------|-----|-----|---|----------------------|-------------------|
| 28 | 59.5 | 4.4 | 178 | 3 | US-09-220-731-24 | Sequence 24, Appl |
| 29 | 59.5 | 4.4 | 178 | 4 | US-09-242-999-24 | Sequence 24, Appl |
| 30 | 59.5 | 4.4 | 144 | 4 | US-09-252-991A-25578 | Sequence 25578, A |
| 31 | 59 | 4.4 | 171 | 4 | US-09-107-532A-3979 | Sequence 3979, Ap |
| 32 | 59 | 4.4 | 180 | 4 | US-09-194-905-5 | Sequence 5, Appl |
| 33 | 59 | 4.4 | 199 | 4 | US-09-252-991A-30363 | Sequence 30363, A |
| 34 | 58.5 | 4.3 | 167 | 4 | US-09-328-352-6278 | Sequence 6278, Ap |
| 35 | 58.5 | 4.3 | 198 | 3 | US-09-413-814-87 | Sequence 87, Appl |
| 36 | 58.5 | 4.3 | 203 | 4 | US-09-252-991A-24921 | Sequence 24921, A |
| 37 | 58 | 4.3 | 136 | 4 | US-09-252-991A-30622 | Sequence 30622, A |
| 38 | 58 | 4.3 | 168 | 1 | US-08-451-947-6 | Sequence 6, Appl |
| 39 | 58 | 4.3 | 168 | 2 | US-08-424-826A-6 | Sequence 6, Appl |
| 40 | 58 | 4.3 | 168 | 3 | US-08-928-694-6 | Sequence 6, Appl |
| 41 | 58 | 4.3 | 168 | 4 | US-08-450-842-6 | Sequence 6, Appl |
| 42 | 58 | 4.3 | 168 | 4 | US-08-451-390-6 | Sequence 6, Appl |
| 43 | 58 | 4.3 | 168 | 5 | PCT-US91-06950-6 | Sequence 6, Appl |
| 44 | 58 | 4.3 | 183 | 4 | US-09-252-991A-31347 | Sequence 31347, A |
| 45 | 58 | 4.3 | 187 | 4 | US-09-252-991A-29325 | Sequence 29325, A |

RESULT 1
US-08-858-207A-519
Sequence 519, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: NO. 6348328e1 Compounds
NUMBER OF SEQUENCES: 352
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858, 207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 519:
SEQUENCE CHARACTERISTICS:
LENGTH: 170 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-519
Query Match 5.6%; Score 75.5; DB 4; Length 170;
Best Local Similarity 22.1%; Pred. No. 2.2;

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/033,003
FILING DATE: December 16, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sawislak, Deborah A
REGISTRATION NUMBER: 37,438
REFERENCE/DOCKET NUMBER: 96-41
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6672
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-991-890-4

Query Match 5.3%; Score 71; DB 3; Length 159;
Best Local Similarity 23.8%; Pred. No. 6.3;
Matches 34; Conservative 18; Mismatches 43; Indels 48; Gaps 7;

49 VILGAVLSRIDAGE-----OLGRIHNSON-----DVEYSPVTEKH 87
DB 12 LLLCGAFA--VSPSEIHAEFGRRHHHGGSGAELRGCGPRFGKHLSTYCPMPKRT 69
OY 88 LTPDMVRELCSAATMSDNTANLLTTIGPKELTAFLEHNMGDHVTALDRWPELENEA 147
DB 70 FT-----TPGCMLES-GRKENVSTSNKKGAL-----GTSEF 105
OY 148 IPNDERDTTVVANA-TTLRLKL 169
DB 106 IPNLSPELKKPLSEGOPLSKII 128

RESULT 5
US-09-328-352-4491
Sequence 4491, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 4491
LENGTH: 169
TYPE: PRT
ORGANISM: *Acinetobacter baumannii*
US-09-328-352-4491

Query Match 5.1%; Score 69; DB 4; Length 169;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 36; Conservative 34; Mismatches 60; Indels 32; Gaps 7;

OY 118 GGPKELEAFLEHNMGDHVTALDRWPELENEAIPNDERDTT-----TPVAMATTLRLKL 169
DB 5 GGEVIMKMLAQLO--RFQWVEQIVYL--DRLTVRERIMVFTTIFVVIVIGYS 58
OY 170 TGELETLASROO-----LIDNMEADYVAGLALSALPAGFTADKSGAGER--GSRG 219
DB 59 LMKHSHSLAEQOQKRLNDKLDKLVMMQSNV-----TMKPALELELDKSGRIQRYAQQG 112
OY 220 IIALGPGKPSRIVITYTGSQATDERNRROIAGSLIK 261
DB 113 LTVSSQONGEQIYV--THQNTAILANFLTLQMLGSLIOK 152

RESULT 6
US-09-010-809-19
Sequence 19, Application US/09010809B
Patent No. 6090601
GENERAL INFORMATION:
APPLICANT: Gustafsson, Claes
APPLICANT: Bellach, Mary C.
TITLE OF INVENTION: Epithelone Polyketide Synthases and Encoding DNA
FILE REFERENCE: 30062-20020.00
CURRENT APPLICATION NUMBER: US/09/010,809B
CURRENT FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 19
LENGTH: 158
TYPE: PRT
ORGANISM: *Sorangium cellulosum*
US-09-010-809-19

Query Match 4.9%; Score 65.5; DB 3; Length 158;
Best Local Similarity 27.6%; Pred. No. 26;
Matches 37; Conservative 18; Mismatches 48; Indels 31; Gaps 6;

OY 106 DNTANLLT-----TIGPKELTAFLEHNMGDHVT--RLDRWPE-----ELNEAIPNDERD 154
DB 21 NHDARKHLITSRGASAPGADVLRSELALGASVTLACVADPRALKDLNDIP----- 75
OY 155 TTPVAMATTLRLKLIGELTTLASROQLIDMADYVAGLRSR-----LPAGCF 205
DB 76 SAHFVAAVYHNASVDGDLGMSLERT-----DRVFAKIDAMWHLHOTODKPLAALF 129
OY 206 IADKSGAGERSRG 219
DB 130 ILFSSVAGVLSSG 143

RESULT 7
US-09-239-909-2
Sequence 2, Application US/09239909
Patent No. 6284952
GENERAL INFORMATION:
APPLICANT: Kumho Petrochemical Co. Ltd.
TITLE OF INVENTION: Transgenic Plants with Divergent SCAM4 or SCAM5 Gene to Achieve
FILE REFERENCE: P99-2-6
CURRENT APPLICATION NUMBER: US/09/239,909
CURRENT FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: EP 99300136.1
EARLIER FILING DATE: 1999-01-08
NUMBER OF SEQ ID NOS: 4
SOFTWARE: KOPATIN 1.0
SEQ ID NO 2
LENGTH: 150
TYPE: PRT
ORGANISM: *G. max calmodulin4 (SCAM4)*
US-09-239-909-2

Query Match 4.8%; Score 65; DB 3; Length 150;
Best Local Similarity 27.9%; Pred. No. 27;
Matches 36; Conservative 17; Mismatches 60; Indels 16; Gaps 6;

OY 73 SONLVESPYTEKRLTDS--MTVRELCSAATMSDNTANLLTIT-----GGPKE 122
DB 6 SEEQIVDKKEAPGLFDKDGDCITVEELATVIRSLQNPTEELQDMISEVDADGNGTIE 65
OY 123 LTFALHNMGDHVTALDRWPELENEAIPNDERDTTVVAMATTLRLKL--TGELETLASRO 180
DB 66 FDEFLSLAKKRYKTDIA--EELAEAFKVFYDKDQNGTIS--ASELRHWININGEKLDEEVE 123
OY 181 QLIDMMEAD 189
DB 124 QMI--READ 130


```

APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESS: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/445,307
FILING DATE: 07 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: CRAIG, ANNE I.
REGISTRATION NUMBER: 32,976
REFERENCE/DOCKET NUMBER: 017,605
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 466-6000
TELEFAX: (617) 466-6040
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-296-195

Query Match      4.8%; Score 65; DB 3; Length 181;
Best Local Similarity 20.5%; Pred. No. 36;
Matches 35; Conservative 24; Mismatches 66; Indels 46; Gaps 5;

QY      4 TLVVKDAEQLGARGVY--TELDNSGELISFREERPPMSTRVYLICGAVLSRIDA 61
DB      17 THAIRVVAATESYLAIRNTSLDSEDELVDYIHNNGVQESYRYV----- 65
QY      62 GGEOLGRIRIYSQ-----NDLVEYSPYTERKHLTDGMTVRELCSAATMSDNTANILL 114
DB      66 ARQSCRPNAAQLEAVFEANONTTAKIETIRASIDGLEV----- 105
QY      115 TTIGPRELTAFLHNMGDHVTRLDRWEPELNEAIPND---ERDTTPVAM 161
DB      106 --IIGINDLDAFRHYDGRITIIQRDNGYSINGNAPREIDLRKRTVTPIRM 154

RESULT 11
US-09-252-991A-29942
Sequence 29942, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 29942
```

```

LENGTH: 178
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: UNSURE
LOCATION: (116)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-29942

Query Match      4.8%; Score 64.5; DB 4; Length 178;
Best Local Similarity 26.8%; Pred. No. 40;
Matches 37; Conservative 14; Mismatches 48; Indels 39; Gaps 8;

QY      110 ANLLTTIG---GPELTAFA---LHNMGDHVTRLDRWEPELNEAIPNDRDTPPVAMA 162
DB      25 ARVLLVEIAQQRRPGEIAAHAGIAHHHPGDHVAR-----AGEDLP---RDVQGLGAAA 74
QY      163 TTLKRLTGLL-----TLASRQQLIDWMEADVYAGPLNSALPAGWFIADKSG 211
DB      75 LRHHLYLAGALQGEHHEBGIGDGTAAAGEQAVVG--ODQEVAGPXY--GIQAGLFLMWQGD 130
QY      212 AGERSGRIIAALGPDKR 229
DB      131 A-----LVGVQQAQR 141

RESULT 12
US-08-624-677A-2
Sequence 2, Application US/08624677A
Patent No. 6476192
GENERAL INFORMATION:
APPLICANT: Lally, Nicola C.
APPLICANT: Jenkins, Mark C.
APPLICANT: Dubey, Jitender P.
TITLE OF INVENTION: Antigenic Useful for the Serodiagnosis of
TITLE OF INVENTION: Neosporosis
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Graeter, Janelle S.
STREET: Rm. 411, Bldg. 005, BARC-W
CITY: Beltsville
STATE: MD
COUNTRY: USA
ZIP: 20705
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,677A
FILING DATE: 15-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Graeter, Janelle S.
REGISTRATION NUMBER: 35,024
REFERENCE/DOCKET NUMBER: 0228.95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-504-5060
TELEFAX: 301-504-6629
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 203 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-624-677A-2

Query Match      4.7%; Score 64; DB 4; Length 203;
Best Local Similarity 23.2%; Pred. No. 55;
Matches 38; Conservative 22; Mismatches 74; Indels 30; Gaps 6;

QY      54 AVLSRIDAGQEQGRIRIHSQNDLVEYSPYTEKHILTDGMTVRELCSAA--ITMSDNTAAN 111
```

Db 33 AGVSATDGDGDAAGNVD-----SDVTDAITGEMPRVSGKPHHTQKSLRK 83
Qy 112 LLLTTIGGPKETAFIHNMGDHY-TRLDREPEPELNEAIPNDERDT-----T 157
Db 84 LAVPVVGA---LTSYL--VADRVLPELTSABEGTESIPGKRKRTNVIAGIAALVAAAAFA 138
Qy 158 PVAAATTLRKLLTGELITLASROOLIDMEADKVA GPILRSALP 201
Db 139 GIGLARTFRHFVPRKSKTVASEDSALCNSEEQYVEGVNGSSDP 182

RESULT 13
US-08-225-480-4
; Sequence 4, Application US/08225480
; Patent No. 581793
; GENERAL INFORMATION:
; APPLICANT: LEVY, STUART B.
; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,480
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,085
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: T0359/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-225-480-4

Query Match 4.7%; Score 63.5; DB 2; Length 144;
Best Local Similarity 29.5%; Pred. No. 37;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

Qy 66 LGRRIHY--SQND--LVEY--SP--VTEKHLTDGNTVRELCSAITMSDNTAANLLTTIG 118
Db 14 LGRLLHVNQKRDRLINELSPDITAAQFVLCISIR--CAACIT-----56
Qy 119 GKRLTAFLH-NMGDHYTRLDRE-----WEPELNEAIPNDERD-----TTTPVAAATTLRK 167
Db 57 -PVELKRVLSVDGALTRMDRLVCKGWVERLPN--PNDKRGVLVLTGGAICGCHQ 113
Qy 168 LLTGELLTASROOLIDMEADKVA--GPILRSALP 201
Db 114 LVGQDL-----HDELTKNLTADEVATLEYLLKRVLP 144

RESULT 14

US-09-118-445-4
; Sequence 4, Application US/09118445
; Patent No. 6391545
; GENERAL INFORMATION:
; APPLICANT: LEVY, STUART B.
; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,445
; FILING DATE: 17-Jul-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/225,480
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: T0359/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-118-445-4

Query Match 4.7%; Score 63.5; DB 4; Length 144;
Best Local Similarity 29.5%; Pred. No. 37;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

Qy 66 LGRRIHY--SQND--LVEY--SP--VTEKHLTDGNTVRELCSAITMSDNTAANLLTTIG 118
Db 14 LGRLLHVNQKRDRLINELSPDITAAQFVLCISIR--CAACIT-----56
Qy 119 GKRLTAFLH-NMGDHYTRLDRE-----WEPELNEAIPNDERD-----TTTPVAAATTLRK 167
Db 57 -PVELKRVLSVDGALTRMDRLVCKGWVERLPN--PNDKRGVLVLTGGAICGCHQ 113
Qy 168 LLTGELLTASROOLIDMEADKVA--GPILRSALP 201
Db 114 LVGQDL-----HDELTKNLTADEVATLEYLLKRVLP 144

RESULT 15
US-08-557-122A-6
; Sequence 6, Application US/08557122A
; Patent No. 5879664
; GENERAL INFORMATION:
; APPLICANT: Hjort, Carsten Møllend
; TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 5879664 of No. 5879664 of No. 5879664 of America, Inc.
; STREET: 405 Lexington Avenue, 64th floor
; CITY: New York
; STATE: New York

Thu Sep 11 16:07:56 2003

seq2_30e_157t.raii

Page 7

```

1 COUNTRY: United States of America
2 ZIP: 10174-6401
3
4 COMPUTER READABLE FORM:
5
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM PC compatible
8 OPERATING SYSTEM: PC-DOS/MS-DOS
9 SOFTWARE: Patentin Release #1.0, Version #1.30
10
11 CURRENT APPLICATION DATA:
12
13 APPLICATION NUMBER: US/08/557,122A
14 FILING DATE: 11-DEC-1995
15 CLASSIFICATION: 435
16
17 ATTORNEY/AGENT INFORMATION:
18
19 NAME: Lambiris, Elias J.
20 REGISTRATION NUMBER: 33,728
21 REFERENCE/DOCKET NUMBER: 3980.204-US
22 TELECOMMUNICATION INFORMATION:
23
24 TELEPHONE: 212-867-0123
25 TELEFAX: 212-878-9655
26
27 INFORMATION FOR SEQ ID NO: 6:
28
29 SEQUENCE CHARACTERISTICS:
30
31 LENGTH: 174 amino acids
32 TYPE: amino acid
33 TOPOLOGY: linear
34
35 MOLECULE TYPE: protein
36
37 US-08-557-122A-6

```

```

Query Match 4.7% Score 63; DB 2; Length 174;
Best Local Similarity 23.98; Pred. No. 56;
Matches 28; Conservative 22; Mismatches 51; Indels 16; Gaps 4;

QY 7 RKVDAEDQGAR-VGYIELDNLSGEIISFRRERPMSTERYLGGAVISRDAQEQ 65
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 KKEGAATLKKENIPLVKVDCETEEBALCRDQCVGGPTLKIRFGL-----DAKVPY 117
   : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 66 LGR-----IHS-QNDLVEYSPTYEKHLTDGMTRELCSAAITMSDNTANLLIT 116
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 118 QGARGTEATVSYWVKOSLPAVSPYTPENLEELKTKMDKIYVIGIASDDQTAINDFTT 174

```

Search completed: September 10, 2003, 12:31:05
Job time : 12.8571 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds
(without alignments)
2183.941 Million cell updates/sec

Title: SEQ2_30E_157T
Perfect score: 1348
Sequence: 1 HPEPLVAVKDAEDQLGARVG.....TMDERNRQIAETIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications, AA.*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubppaa/PC1_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubppaa/PC1US_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 457 | 33.9 | 94 | US-10-102-806-760 | Sequence 760, App |
| 2 | 75 | 5.6 | 202 | US-10-127-816-9 | Sequence 9, Appl |
| 3 | 75 | 5.6 | 202 | US-10-142-717-12 | Sequence 12, Appl |
| 4 | 70.5 | 5.2 | 193 | US-09-805-354-8 | Sequence 8, Appl |
| 5 | 70.5 | 5.2 | 193 | US-10-144-359-8 | Sequence 8, Appl |
| 6 | 70 | 5.2 | 206 | US-10-156-761-13867 | Sequence 765, App |
| 7 | 69.5 | 5.1 | 149 | US-10-101-464A-765 | Sequence 8136, App |
| 8 | 69 | 5.0 | 149 | US-10-156-761-8136 | Sequence 8136, App |
| 9 | 67.5 | 5.0 | 187 | US-09-348-018-8 | Sequence 8, Appl |
| 10 | 67.5 | 5.0 | 178 | US-09-738-626-5680 | Sequence 5680, App |
| 11 | 67 | 5.0 | 194 | US-10-156-761-8623 | Sequence 8623, App |
| 12 | 66 | 4.9 | 202 | US-10-127-816-11 | Sequence 11, Appl |
| 13 | 65.5 | 4.9 | 191 | US-09-764-870-390 | Sequence 390, App |
| 14 | 65.5 | 4.9 | 191 | US-10-125-540-390 | Sequence 390, App |
| 15 | 65.5 | 4.9 | 191 | US-10-125-540-390 | Sequence 390, App |

| | | | | | | |
|----|------|-----|-----|----|---------------------|--------------------|
| 16 | 64.5 | 4.8 | 189 | 9 | US-09-815-242-5862 | Sequence 5862, App |
| 17 | 64.5 | 4.8 | 189 | 9 | US-09-815-242-12979 | Sequence 12979, A |
| 18 | 64.5 | 4.8 | 191 | 15 | US-10-156-761-8978 | Sequence 8978, App |
| 19 | 64 | 4.7 | 162 | 10 | US-09-738-626-4796 | Sequence 4796, App |
| 20 | 64 | 4.7 | 179 | 15 | US-09-764-868-757 | Sequence 757, App |
| 21 | 64 | 4.7 | 179 | 15 | US-10-106-698-4858 | Sequence 4858, App |
| 22 | 64 | 4.7 | 202 | 15 | US-10-189-346-12 | Sequence 12, Appl |
| 23 | 63.5 | 4.7 | 88 | 15 | US-10-100-252-6 | Sequence 6, Appl |
| 24 | 63.5 | 4.7 | 144 | 15 | US-10-131-406-4 | Sequence 4, Appl |
| 25 | 63.5 | 4.7 | 206 | 15 | US-09-738-626-5425 | Sequence 5425, App |
| 26 | 62.5 | 4.6 | 190 | 15 | US-10-156-761-9507 | Sequence 9507, App |
| 27 | 62.5 | 4.6 | 195 | 15 | US-10-156-761-12656 | Sequence 12656, A |
| 28 | 62 | 4.6 | 174 | 9 | US-09-864-761-35777 | Sequence 35777, A |
| 29 | 62 | 4.6 | 191 | 11 | US-09-927-827-461 | Sequence 64, Appl |
| 30 | 62 | 4.6 | 202 | 15 | US-10-156-761-13460 | Sequence 13460, A |
| 31 | 61.5 | 4.6 | 142 | 10 | US-09-862-027-31 | Sequence 31, Appl |
| 32 | 61.5 | 4.6 | 196 | 10 | US-09-738-626-6624 | Sequence 6624, App |
| 33 | 61.5 | 4.6 | 200 | 15 | US-10-189-346-16 | Sequence 12086, A |
| 34 | 61 | 4.5 | 189 | 9 | US-10-101-464A-616 | Sequence 616, App |
| 35 | 61 | 4.5 | 189 | 15 | US-09-764-870-523 | Sequence 4637, App |
| 36 | 61 | 4.5 | 197 | 15 | US-10-125-540-523 | Sequence 523, App |
| 37 | 61 | 4.5 | 197 | 10 | US-09-803-661-5 | Sequence 5, Appl |
| 38 | 60.5 | 4.5 | 197 | 15 | US-10-300-827-5 | Sequence 230, App |
| 39 | 60.5 | 4.5 | 160 | 12 | US-09-882-227-230 | Sequence 16, Appl |
| 40 | 60 | 4.5 | 202 | 15 | US-10-189-346-16 | Sequence 616, App |
| 41 | 60 | 4.5 | 185 | 15 | US-10-101-464A-616 | Sequence 4637, App |
| 42 | 59.5 | 4.4 | 190 | 10 | US-09-738-626-4637 | Sequence 12056, A |
| 43 | 59.5 | 4.4 | 182 | 15 | US-10-156-761-12056 | Sequence 265, App |
| 44 | 59.5 | 4.4 | 196 | 15 | US-10-219-220-265 | Sequence 461, App |
| 45 | 59.5 | 4.4 | 206 | 9 | US-09-741-669-461 | Sequence 10335, A |
| | | | | 9 | US-09-815-242-10335 | |

ALIGNMENTS

RESULT 1
US-10-102-806-760
; Sequence 760, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: Nucleic acids, proteins and antibodies
; FILE REFERENCE: P103P1C1
; CURRENT FILING DATE: 2002-03-22
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 760
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (91)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760

Query Match 33.9%; Score 457; DB 15; Length 94;
Best Local Similarity 97.8%; Pred. No. 1.3e-39;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 75 NDVLYSPYTERKHLTGOMYRELCSAIIYMSDNTAAILTTIGGPKELTAFLAHMGDHY 134
DB 1 NDVLYSPYTERKHLTGOMYRELCSAIIYMSDNTAAILTTIGGPKELTAFLAHMGDHY 60

QY 135 TRLDWEPELNEAIPNDERDTTPVAMATT 164
|||
Db 61 TRLDWEPELNEAIPNDERDTTPVAMATT 90

RESULT 2

US-10-127-816-9
; Sequence 9, Application US/10127816
; Publication No. US20030104616A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Fox, Brian A.
; APPLICANT: Klucher, Kevin M.
; APPLICANT: Taft, David W.
; APPLICANT: Kindsvogel, Wayne R.
; TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
; FILE REFERENCE: 01-17
; CURRENT APPLICATION NUMBER: US/10/127,816
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,408
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/286,482
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/341,050
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/341,105
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 09/895,834
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/285,424
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-816-9

Query Match
Best Local Similarity 5.6%; Score 75; DB 15; Length 202;
Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

QY 50 LCGAVLSRIDAOGEOLGRRIHYSQND--LYEYSVTREKHL-----TDGMYREL----- 97
|||
Db 17 LLLAAVLTROQADVPVPRATRLPVEAKDCHIAQFSLSPKELOAFKAKDAIEKRILEKDL 76
QY 98 -CSA-----ATMS--DNTANLLITIGGPKELTAFLELN 129
|||
Db 77 RCSSHLFPRAWMDLKOLOVOERPKALAEVALTLTKVEMMTSALATITLGGPLHTLSHHS 136
QY 130 MGDHYT-----RLDWEPELNEAIPNDERDTTPVAMATTLRKILTGSELTL 176
Db 137 OLQCTQLOATAEPSPSRRLSRMLHRLQEA-QSKETPGCEASVTSNLFRLITDLCV 195
QY 177 ASRQOLI 183
Db 196 ANGDOCV 202

RESULT 3

US-10-142-717-12
; Sequence 12, Application US/10142717
; Publication No. US20030104579A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Baum, Peter R.
; APPLICANT: Mosley, Bruce A.
; APPLICANT: Ketchum, Randal R.
; APPLICANT: Taylor, Scott L.
; TITLE OF INVENTION: CYTOKINE POLYPEPTIDES
; FILE REFERENCE: 3282-A

; CURRENT APPLICATION NUMBER: US/10/142,717
; CURRENT FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.1
; SEQ ID NO 12
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-142-717-12

Query Match
Best Local Similarity 5.6%; Score 75; DB 15; Length 202;
Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

QY 50 LCGAVLSRIDAOGEOLGRRIHYSQND--LYEYSVTREKHL-----TDGMYREL----- 97
|||
Db 17 LLLAAVLTROQADVPVPRATRLPVEAKDCHIAQFSLSPKELOAFKAKDAIEKRILEKDL 76
QY 98 -CSA-----ATMS--DNTANLLITIGGPKELTAFLELN 129
|||
Db 77 RCSSHLFPRAWMDLKOLOVOERPKALAEVALTLTKVEMMTSALATITLGGPLHTLSHHS 136
QY 130 MGDHYT-----RLDWEPELNEAIPNDERDTTPVAMATTLRKILTGSELTL 176
Db 137 OLQCTQLOATAEPSPSRRLSRMLHRLQEA-QSKETPGCEASVTSNLFRLITDLCV 195
QY 177 ASRQOLI 183
Db 196 ANGDOCV 202

RESULT 4

US-09-805-354-8
; Sequence 8, Application US/09805354
; Publication No. US20030078375A1
; GENERAL INFORMATION:
; APPLICANT: Arnaout, M. Amln
; APPLICANT: Li, Rui
; APPLICANT: Xiong, Jian-Ping
; TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 00786-536001
; CURRENT APPLICATION NUMBER: US/09/805,354
; CURRENT FILING DATE: 2002-06-04
; PRIOR APPLICATION NUMBER: US 09/758,493
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/221,950
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-805-354-8

Query Match
Best Local Similarity 5.2%; Score 70.5; DB 11; Length 193;
Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;

QY 59 IDAGOEOGLRRIRHYSQNDLYESP--VTEKHLTDGTVRELCSAATMSDNTANLLIT 116
|||
Db 36 IGFGIOIYG-----VOYGEDVVEHFLNDYASVDVNAASHIEDR----- 77
QY 117 ICGPKELTAPLHNNGDHYTRLDRWE-----PELNEAIPNDER 154
|||
Db 78 -GGTETRTAF-----GIEPARSEAFQKGRKAKKVIITDGSNDSPLEKYIQGSERD 132
QY 155 TTPPVAMA 162
Db 133 NTRYAYVA 140

RESULT 5


```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8136
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Streptomyces avermectilis
; US-10-156-761-8136

Query Match
Best Local Similarity 30.9%; Pred. No. 35;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;

QY 168 LITGELTL-----ASROQLIDMMEADKVAGPLLRSGALPAGWFIADKSG----- 211
DB 50 LLDIELTLTKRLILVAVSYDAKEMCIDWMD-----PALSSHADGRELAEENRRRLREEL 105

QY 212 AGERSRGITIALGPDGKPSR 232
DB 106 AGLRDADALPSAGSPAEKPER 126

RESULT 9
US-10-233-926-4
; Sequence 4, Application US/10233926
; Publication No. US20030131382A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLYLTRANSFERASE
; FILE REFERENCE: BBI419 US NA
; CURRENT APPLICATION NUMBER: US/10/233,926
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: US/09/735,846
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
; US-10-233-926-4

Query Match
Best Local Similarity 21.4%; Pred. No. 50;
Matches 22; Conservative 23; Mismatches 33; Indels 25; Gaps 4;

QY 159 VAMATTTLRKLL-----TGGELLTLASRQQLIDMMEADKVAGPLLRSGALPAGWFIADKSGA 212
DB 14 LSLXSJLPLNLAMADHAAPQSSQEEEDMKREAGGADV-----EVAADRGGG 64

QY 213 GERSRGITIALGPDGKPSRIV-----IYTGSOQATMDERNR 250
DB 65 GGAANGCI-----PGCRPIRYADGIYDLFHFHGAASLEQAKR 102

RESULT 10
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: Thell et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF
```

```

; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-948-018-8

Query Match
Best Local Similarity 22.4%; Pred. No. 69;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

QY 112 LLTTTGGPKRELATLHNGDHTVRLDRWEPE-----LNEALPMDERDTTPVAMAT- 163
DB 11 LLLALAPPPEAS-----QYCGRLRYWNPDKCCSSCLQRGP-----PCTGALETG 58

QY 164 -----TLRKLLTGGELTLASR--COLIDMMEA-----DRVAGP-----L 195
DB 59 DTWKASLLPLLSRELSSLASQPLSKRLDELFLVLELIVLDPDPGGGMAHGTTHLA 118

QY 196 LRSALPAGWFIADKSGAGERSRGITIALGPDGKPSR--IVITYTGSQATMDERNRQIA 253
DB 119 ARYGLPAAW-----STFAYSLRPSRSPLALILEMVAARPSASIGOLGTHLA 165

QY 254 ETG 256
DB 166 QLG 168

RESULT 11
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAMA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-5680

Query Match
Best Local Similarity 24.8%; Pred. No. 73;
Matches 25; Conservative 17; Mismatches 37; Indels 22; Gaps 4;

QY 171 GELLTLASRQQLIDW-----MEADKVAGPLLRSGALPAGWFIADKSGAGERSRGITIA 223
```

```

Db      61 GELFDALAEQRLNGAGVLTLEVSIPGVDPNPL--LTPAHM-----RRNRGLVA 107
Qy      224 LGPDGKP--SRIWVYTTGSGATMDERNQIAEIGASLKH 262
          + ||| :|| : : |||::: | :
          108 LDGDKRRVARIGALNDAETHVLLERNKKLLEYVTLLEAH 148

```

```

RESULT 12
US-10-156-761-8623
; Sequence 8623, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMCRA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8623
; LENGTH: 194
; TYPE: PR1
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

```

| | 5.08; | Score 67; | DB 15; | Length 194; |
|-----------------------|-------------|--|------------------------|-------------|
| Query Match | 38.68; | Pred. No. 82; | | |
| Best Local Similarity | Matches 17; | Conservative 9; | Mismatches 12; | Indels 6; |
| | | | | Gaps 2 |
| QY | 202 | AGMFLADKSGAGGGRSG--- | IIAALGPDG--KKSRIYIVYTT | 239 |
| | : | : : : : : : | : : : : : : | : |
| Db | 75 | SGFRYTPRSGAGSERVTGCAEYRITAAIGFLAVHEPRVYAAVYT | | 118 |

```

: RESULT 13
: US-10-127-816-11
: Sequence 11, Application US/10127816
: Publication No. US20030104416A1
: GENERAL INFORMATION:
:   APPLICANT: Sheppard, Paul O.
:   APPLICANT: Fox, Brian A.
:   APPLICANT: Kluehner, Kevin M.
:   APPLICANT: Taft, David W.
:   TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
:   FILE REFERENCE: 01-17
:   CURRENT APPLICATION NUMBER: US/10/127, 816
:   CURRENT FILING DATE: 2002-04-19
:   PRIOR APPLICATION NUMBER: US 60/285,408
:   PRIOR FILING DATE: 2001-04-20
:   PRIOR APPLICATION NUMBER: US 60/286,482
:   PRIOR FILING DATE: 2001-04-25
:   PRIOR APPLICATION NUMBER: US 60/341,050
:   PRIOR FILING DATE: 2001-10-22
:   PRIOR APPLICATION NUMBER: US 60/341,105
:   PRIOR FILING DATE: 2001-10-22
:   PRIOR APPLICATION NUMBER: US 09/895,834
:   PRIOR FILING DATE: 2001-06-29
:   PRIOR APPLICATION NUMBER: US 60/285,424
:   NUMBER OF SEQ ID NOS: 59
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 11
:   LENGTH: 202

```

! TYPE: PRT
! ORGANISM: Homo sapiens
US-10-127-816-11

| | | | | |
|-----------------------|-----------------|-------------------|-----------|------------|
| Query Match | 4.98 | Score 66 | DB 15 | Length 202 |
| Best Local Similarity | 21.58 | Pred. No. 1.1e+02 | | |
| Matches 41 | Conservative 23 | Mismatches 65 | Indels 62 | Gaps 8 |

```

QY 50 LIGGAVSLRIDAGEQGLRTHYSOND--LVEYSPTYEKHL-----TDGMTVREL----- 97
Db 17 LILAAVLTTRIOADPPVPVPRATRLPEVKAEDCHIAQFKSLSPKELQAFKKAQATERLLEKDM 76
QY 98 -CSA-----AATMS--DNTAAAILLTITGGPKELTATLHN 129
Db 77 RGSNHLISRAMDLKOLQOYERKPAQAEVATLTKWENINDSALTTTLGPPHLTLSHIRS 136
QY 130 MGHVHT-----RLDRWEPELNEA---IPNDERDPTTPVAAAMATTLERLGTGE 172
Db 137 QLOQTQLOQATAEPKPPSRRLSRWLRHLRQEAQSKETPGGLEDSVT-----SNLFQLLNRD 191
QY 173 LILTLASRQOLI 183
Db 192 LKCVASGDCVC 202

```

```

US-09-764-870-390
US-09-764-870-390
; RESULT 14
; Sequence 390, Application US/09/764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ14
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-870-390

```

| | | | | |
|--------------------------|--------|--------------------|------------|-------------|
| Query Match | 4.98; | Score 65.5; | DB 9; | Length 191; |
| Best Local Similarity | 19.78; | Pred. No. 1.1e+02; | | |
| Matches 34; Conservative | 20; | Mismatches 48; | Indels 71. | Gaps 7. |

QY 21 YIELDNGCELLIESRPERFERPMSTFFKVLICGAVLSRIAGOBQLGRIRHYSQNDLVEY 80
Db | | | :
48 YCRSLTNSSEV-----DFTGDYDFSETENSEIAQVDLEAK-----YMO----- 85
QY 81 SPVTEKHLDGTARELCSAITMSDNLTANLTTTGCPRELTAFPL- 128
| | | | : | | | | :
Db 86 LPEKKKHTDPAT-ENLCSESI-----KNKLSTTTGNLTLOTXKHTENOSGEGEVTI 138
QY 129 -----MGDHVTRLDSEPE-----LNPAIPND 151
| | | : | | | :
Db 139 EPGADLLDYVPSSQAIFYKNLONSSMDLCDSHSKKDROMASSNHTVNEELPHN 191

RESULT 15
US-10-125-540-390

```
; Sequence 390, Application US/10125540
; Publication No. US20030059875A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214C1
; CURRENT APPLICATION NUMBER: US/10/125,540
; CURRENT FILING DATE: 2002-04-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 390
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (124)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (157)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-10-125-540-390

Query Match          4.98; Score 65.5; DB 15; Length 191;
Best Local Similarity 19.7%; Pred. No. 1.1e+02;
Matches 34; Conservative 20; Mismatches 48; Indels 71; Gaps 7;

QY 21 YIELDNSGILSFREPERPMWSTFKVLLCGAVLSRIDAGOEOLGRRIHYSNDLVEX 80
   | | | | |
DB 48 YCRLSINDSEV-----DMFGDYDSTFENSFIAOVDLEOK-----YMQ----- 85

QY 81 SPYTEKHLTDGMTVRELCSAATWMSDNTANILLTTIGSPKELTAFLH----- 128
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 86 LPEHKKHATDPAT-ENLCSESI-----KNKLSTITIGNLTLQTKKHTEGSGYEGVTI 138

QY 129 -----NKGDHVTRLDRWEPE-----LNEAIPND 151
   : | | | : | | | : | | | : | | | : | | | : | | | : | | | : | | |
DB 139 EPGADLLYDVPSQAIFYXNLQNSSNDLGDHSMKDRDWXSSSHNTVNEELPHN 191

Search completed: September 10, 2003, 12:33:20
Job time : 18.5714 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 10.2857 Seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_30E_157T

Perfect score: 1348
Sequence: 1 HPEPLVKKVDAEDQIGARVQ.....TMDERNRQIAIGASLIRKM 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | ID | Description |
|------------|-------|---------------|--------|----|-------------|
| 1 | 524 | 38.9 | 105 | 2 | JC2566 |
| 2 | 77 | 5.7 | 191 | 2 | S67447 |
| 3 | 75.5 | 5.6 | 190 | 2 | T09136 |
| 4 | 74.5 | 5.5 | 200 | 2 | G97064 |
| 5 | 74.5 | 5.5 | 113 | 2 | T45195 |
| 6 | 74.5 | 5.5 | 172 | 2 | AD3606 |
| 7 | 73.5 | 5.3 | 195 | 2 | T36975 |
| 8 | 73.5 | 5.3 | 149 | 2 | F71252 |
| 9 | 71.5 | 5.3 | 152 | 2 | T36984 |
| 10 | 71 | 5.3 | 153 | 2 | E81708 |
| 11 | 71 | 5.3 | 177 | 2 | A83753 |
| 12 | 71 | 5.3 | 192 | 2 | G83096 |
| 13 | 70 | 5.2 | 108 | 2 | E84217 |
| 14 | 70 | 5.2 | 170 | 2 | AB0192 |
| 15 | 69.5 | 5.2 | 184 | 2 | T21126 |
| 16 | 68.5 | 5.1 | 145 | 2 | F84251 |
| 17 | 68.5 | 5.1 | 167 | 2 | D83360 |
| 18 | 68.5 | 5.1 | 192 | 2 | A83587 |
| 19 | 67.5 | 5.0 | 131 | 2 | AD2281 |
| 20 | 66.5 | 4.9 | 116 | 2 | C82906 |
| 21 | 66.5 | 4.9 | 177 | 2 | D90227 |
| 22 | 66.5 | 4.9 | 180 | 2 | C71869 |
| 23 | 66.5 | 4.9 | 198 | 2 | D95285 |
| 24 | 65.5 | 4.9 | 198 | 2 | F36978 |
| 25 | 65.5 | 4.9 | 181 | 2 | AB1902 |
| 26 | 65.5 | 4.9 | 195 | 2 | AE0623 |
| 27 | 65.5 | 4.9 | 195 | 2 | C69296 |
| 28 | 65 | 4.8 | 146 | 2 | C72703 |
| 29 | 65 | 4.8 | 148 | 2 | E75283 |

| | | | | | | |
|----|------|-----|-----|---|--------|--------------------|
| 30 | 65 | 4.8 | 150 | 2 | T09585 | calmodulin - soybe |
| 31 | 65 | 4.8 | 160 | 1 | E69186 | conserved hypothet |
| 32 | 65 | 4.8 | 168 | 2 | B75498 | conserved hypothet |
| 33 | 65 | 4.8 | 177 | 1 | B43387 | nonstructural prot |
| 34 | 65 | 4.8 | 177 | 2 | J01931 | nonstructural prot |
| 35 | 65 | 4.8 | 180 | 2 | G70912 | hypothetical prote |
| 36 | 64.5 | 4.8 | 145 | 2 | H75262 | hypothetical prote |
| 37 | 64.5 | 4.8 | 151 | 2 | D81333 | hypothetical prote |
| 38 | 64 | 4.7 | 113 | 2 | D70580 | probable protein-t |
| 39 | 64 | 4.7 | 162 | 2 | AG0769 | hypothetical prote |
| 40 | 64 | 4.7 | 178 | 2 | I40124 | probable acetyltra |
| 41 | 64 | 4.7 | 191 | 2 | E95333 | outer surface prot |
| 42 | 63.5 | 4.7 | 164 | 2 | H82336 | hypothetical prote |
| 43 | 63.5 | 4.7 | 166 | 2 | C90029 | regulator of sigma |
| 44 | 63.5 | 4.7 | 168 | 2 | T20606 | hypothetical prote |
| 45 | 63.5 | 4.7 | 179 | 2 | AB1994 | hypothetical prote |

ALIGNMENTS

RESULT 1
JC2566
b1a protein - Pseudomonas aeruginosa (fragment)
C/Species: Pseudomonas aeruginosa
C/Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996
C/Accession: JC2566
R/West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 148[128], 81-86, 1994
A/Title: Construction of Improved Escherichia-pseudomonas shuttle vectors derived fro
A/Reference number: JC2565
A/Note: due to a typographical error the volume number 148 appears as 128
A/Accession: JC2566
A/Molecule type: DNA
A/Residues: 1-105 <WES>
C/Genetics:
A/Gene: b1a
C/Superfamily: beta-lactamase I

Query Match 38.9% Score 524; DB 2; Length 105;
Best local similarity 100.0%; Pred. No. 4.4e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATTLLKRLTGGELTLASROOLIDMWEADKAGPLRSALPGMFIADKSGGERSRG 219
|||||
Db 2 AMATTLLKRLTGGELTLASROOLIDMWEADKAGPLRSALPGMFIADKSGGERSRG 61
QY 220 IIALGPDGKPSRIWVYITTSQATMDERNRQIAIGASLIRKM 263
|||||
Db 62 IIALGPDGKPSRIWVYITTSQATMDERNRQIAIGASLIRKM 105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C/Accession: T38062; S67447
Kumlien, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, March 1996
A/Reference number: 221766
A/Accession: T38062
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-191 <MC2>
A/Cross-references: EMBL:Z69944; NID:q1217974; PIDN:CAA93808.1; PID:q1217978; GSPDB:G
A/Experimental source: strain 972h-; cosmid c1f12
A/Genes: SPAC1F12.04c
A/Map position: 1
C/Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c
Query Match 5.7%; Score 77; DB 2; Length 191;

| | Best Local Similarity | 22.6% | Pred. No. 36: | |
|----|-----------------------|---|---------------|---------------------------------------|
| | Matches | 43; | Conservative | 22; Mismatches 69; Indels 56; Gaps 7; |
| QY | 104 | MSDNTAANLLLTITGGPKELTAE LHNNGDHYTRIDRMEPELENAIPNDERDTTTPYAMAT | 163 | |
| Db | 1 | MSYMSLNLMLOPSSGIDRIATALLVN---VARLD-----PASRSKSPALVSMLN | 46 | |
| QY | 164 | TLRKLLTGLTLLTASRQQLIDMMEADKVAGGLASALPAGWFIADKSGGGEGRSRIIA | 223 | |
| Db | 47 | EFRC-----TLRLPGLYKLIYVNFERKDDSPETYSMAINIGGYVTE--GIAFLGGKQIISI | 99 | |
| QY | 224 | LGPDKP-----SRIIVYITGSGATMDERNROI----- | 252 | |
| Db | 100 | -----SRPLEDKMLWSSRFWLLDTLLTYQLREKTEDEKEHQDLASNYASLPLCIHMS | 155 | |
| QY | 253 | AEIGASLIKH 262 | | |
| Db | 156 | VENGAGLHKH 165 | | |

RESULT 3
T09136
ADP-ribosylation factor homolog ARL3 - Trypanosoma brucei
N:Alternate names: ADP ribosylation factor 3 homolog
C:Species: Trypanosoma brucei
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Feb-2001
C:Accession: T09136
R:Briggaard, F.; Vedeneva, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.,
Mol. Biochem. Parasitol. 94, 249-264, 1998
A:Title: Conserved organization of genes in trypanosomatids.
A:Reference number: Z16580; MUID:58418771; PMID:9747975
A:Accession: T09136
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190 <R1>
A:Cross-references: EMBL:AF031926; NID:g3452211; PIDN:AAC32774.1; PID:g3452215
A:Experimental source: strain AntAt1
A:Note: small G-protein
C:Genetics:
C:Gene: ARL3
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end, lipoprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)
F:89-94/Region: nucleotide-binding motif B
F:126-129/Region: GTP-binding NKXD motif
F:171-176/Region: nucleotide-binding motif B
E:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

RESULT 4
G97064
spore coat protein COTC [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: G97064
R:Moelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: G97064
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-200 <KRR>
A:Cross-References: GB:AE001437; PIDN:AAK79306.1; PID:q15024270; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1338

| | Query Match | 5.6% | Score 75.5; | DB 2; | Length 200; |
|----|--|--------|---------------|-------|----------------------------------|
| | Best Local Similarity | 23.7%; | Pred. No. 51; | | |
| | Matches | 36; | Conservative | 26; | Mismatches 55; Indels 23; Gaps 7 |
| Oy | 1 HPTLYKVKDAEDLGARVGIYELDLNSELLESFR-PEERFPAMSTFKVLGCGAVLSRI | 59 | | | |
| | : : : : : | : | : | : | : |
| Dd | 10 HP---VKIKPNMQL-AKVIIITQYGDPDELAAISIRLRSRFSMTVP-----QAIAITLN | 59 | | | |
| Oy | 60 DACOGLGRRIHYSONDVEFSPVTEKHLLTDCVTRELCSSAAT--MSDNTAANLLTTTI | 117 | | | |
| | : : | : | : : | : | : |
| Dd | 60 DIGTEELAH-----LEIVGSITRQSLRGSLSVETLKSGLDAYFADHSATVPASAA | 110 | | | |
| Oy | 118 GGPKELTAFLNMGDDEVTRL | 137 | | | |
| | : : : : : | : | : | : | : |
| Dd | 111 GNP-FTAAYIQSKGPITDL | 129 | | | |

```

RESULT 5
T45195
hypothetical protein u1756t [imported] - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 21-Jan-2000 #sequence_rev1s1on 21-Jan-2000
C/accession: T45195
R:Robison, R.
submitted to the EMBL Data Library, September 1994
A/Reference number: Z16911
A/Accession: T45195
A/Status: preliminary; translated from GB/EMBL/DDAJ
A/Molecule type: DNA
A/Residues: 1-113 <KEI>
A/Cross-References: EMBL:U15180; PIDN:AAA62885.1

```

| | | | | |
|-----------------------|--------|---------------|-------|----------------------------------|
| Query Match | 5.58; | Score 74.5; | DB 2; | Length 113; |
| Best Local Similarity | 28.68; | Pred. No. 28; | | |
| Matches | 34; | Conservative | 19; | Mismatches 43; Indels 23; Gaps 8 |

| | | | |
|----|-----|---|-----|
| QY | 130 | MGDHY--TRLDREPELN---EAIPEDEDT---TTPVAMATYLRKLTGELLTLASRQ | 180 |
| | | : : : : : : | |
| DB | 1 | MGSDIGMEREGRWGNTQCPLRVVPEDESDPTLDGRASPEDLIT--NILSTPTSHPPPS | 58 |
| | | : : : : : : | |
| QY | 181 | QLIDMMEA-DRVAGPLL-----RSALPGWFIADKSGAGERSGRTIAALGPPDPSR | 232 |
| | | : : : : : : | |
| DB | 59 | RDDDVPEFDLQGTAVFDTATGDKATMPAVGGI-----GATIRGS-GILASLSFFQOPAR | 112 |
| | | : : : : : : | |

RESULT 6
AD3606
molybdopterin biosynthesis mog protein [imported] - *Brucella melitensis* (strain 16M)
C:Species: *Brucella melitensis*
C:date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C/accession: AD3606
R:Belvecchio, V.G.; Kaparakis, V.; Redkar, R.V.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Haglund, S.; O'Callaghan, D.; Lett
Pric. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:title: The genome sequence of the facultative intracellular pathogen *Brucella melit*
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3606

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 ; Search time 5.57143 Seconds
(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_30E_157T
Perfect score: 1348
Sequence: 1 HPELVKRVNDAEDQIGARVG.....TMDERNQIAEIGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query | Length | DB ID | Description |
|------------|-------|---------|--------|-------|-------------|
| 1 | 77 | 5.7 | 191 | 1 | YDA4_SCHPO |
| 2 | 75 | 5.6 | 201 | 1 | RACG_DICDI |
| 3 | 72 | 5.3 | 149 | 1 | NDK_TREPA |
| 4 | 68.5 | 5.1 | 196 | 1 | RAC2_LOTJA |
| 5 | 67 | 5.0 | 122 | 1 | HG31_THIFE |
| 6 | 66.5 | 4.9 | 116 | 1 | RBRV_UREPA |
| 7 | 65.5 | 4.9 | 198 | 1 | OGLI_ARCFU |
| 8 | 65.5 | 4.9 | 206 | 1 | KTHV_METAC |
| 9 | 65 | 4.8 | 177 | 1 | VNSC_RINDK |
| 10 | 65 | 4.8 | 177 | 1 | VNSC_RINDR |
| 11 | 63.5 | 4.7 | 144 | 1 | MARR_ECOLI |
| 12 | 63.5 | 4.7 | 185 | 1 | YCJC_ECOLI |
| 13 | 63.5 | 4.7 | 196 | 1 | ALRH_BACSU |
| 14 | 63 | 4.7 | 176 | 1 | HSLV_THEMA |
| 15 | 62.5 | 4.6 | 150 | 1 | PDGV_SALTY |
| 16 | 62.5 | 4.6 | 172 | 1 | YDEL_ECOLI |
| 17 | 62.5 | 4.6 | 182 | 1 | PYRE_STRCO |
| 18 | 62.5 | 4.6 | 195 | 1 | TRPF_THEVO |
| 19 | 62.5 | 4.6 | 200 | 1 | RR4_PELNE |
| 20 | 62 | 4.6 | 182 | 1 | IHBB_RAT |
| 21 | 62 | 4.6 | 174 | 1 | YB61_BIFLO |
| 22 | 62 | 4.6 | 186 | 1 | YCBR_SALTY |
| 23 | 62 | 4.6 | 202 | 1 | CONT_ELV |
| 24 | 61.5 | 4.6 | 178 | 1 | HSIV_RALSO |
| 25 | 61 | 4.5 | 200 | 1 | SECR_HUMAN |
| 26 | 61 | 4.5 | 201 | 1 | TARV_CAUCR |
| 27 | 60.5 | 4.5 | 103 | 1 | RS10_NEIGO |
| 28 | 60.5 | 4.5 | 146 | 1 | HBC_RABIT |
| 29 | 60.5 | 4.5 | 156 | 1 | BFR_ACOVI |
| 30 | 60.5 | 4.5 | 160 | 1 | TARV_HELPY |
| 31 | 60.5 | 4.5 | 175 | 1 | Y5B1_AQUAE |
| 32 | 60.5 | 4.5 | 184 | 1 | MLR1_SCHPO |
| 33 | 60 | 4.5 | 148 | 1 | CALM_BLAEM |

| | | | | | |
|----|------|-----|-----|---|------------|
| 34 | 60 | 4.5 | 159 | 1 | GRE4_CHLTE |
| 35 | 60 | 4.5 | 178 | 1 | PYRE_ARCFU |
| 36 | 60 | 4.5 | 178 | 1 | UCRI_ANASP |
| 37 | 60 | 4.5 | 184 | 1 | ARL2_DROME |
| 38 | 59.5 | 4.4 | 103 | 1 | RS10_NEIMA |
| 39 | 59.5 | 4.4 | 135 | 1 | CCRN_PAROL |
| 40 | 59.5 | 4.4 | 140 | 1 | YU33_YERPE |
| 41 | 59.5 | 4.4 | 184 | 1 | HRPL_PSESY |
| 42 | 59.5 | 4.4 | 196 | 1 | I196_ASFB7 |
| 43 | 59.5 | 4.4 | 198 | 1 | UPP_ARATH |
| 44 | 59.5 | 4.4 | 205 | 1 | RS4_ECOLI |
| 45 | 59.5 | 4.4 | 206 | 1 | RS4_SHITL |

ALIGNMENTS

RESULT 1
YDA4_SCHPO STANDARD: PRT: 191 AA.
AC YDA4_SCHPO
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein Clf12.04c in chromosome I.
GN SPAC1F12.04c.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
CX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream K.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles S., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltham T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hiderago J., Hodson A.,
RA James K., Jones L., Jones M., Leach S., McDonald S., McLean J.,
RA Mooney P., Mouton S., Mungall K., Murphy L., Niblett D., Odeh C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Slimmons M., Squares R., Stevens S., Taylor K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grympey B.,
RA Welljens I., Vansteleels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesli D., Mueller-Auer S.,
RA Borzym K., Langermann W., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Carrutti L., Love T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT *The genome sequence of Schizosaccharomyces pombe*;
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: Z69944; CAAG3808.1; -
DR PIR: T38062; S67447.
DR GeneDB_Spombe; SPAC1F12.04c; -

Hypothetical protein.
 SQ SEQUENCE 191 AA; 21549 MW; 65555347F0EBED16 CRC64;

Query Match 5.7%; Score 77; DB 1; Length 191;
 Best Local Similarity 22.6%; Pred. No. 26;
 Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;

DB 104 MSDNTANLLTTIGGPKELTAFLHNMGDVTRLDREPELLEAIPNDEROTTTPYAMAT 163
 1 MSYMSLNLMLQNPSSGIDKIAALVYV---VARLD-----PASSKSTAOVLSMLN 46
 OY 164 TLARKLLTGLLELTASLQQLDWMHADKAVAGFLRSALPAGFIADKSGCGERGIIIA 223
 DB 47 EPRC-----TLRLPLGLKLVNFRKSSPPTYSMSAINIGYVTE--GLAFLGKQIISI 99
 OY 224 LAPPDKP-----SRIVYITGTSQATNDENRQI----- 252
 DB 100 ----SKPLEDKMLMSRFWLLDTLTLITQLRKTEDKEHQDLQSLASLPLCIHMS 155
 OY 253 AEIGASLIRH 262
 DB 156 VENGAGLIRH 165

RESULT 2
 RACG_DICDI
 ID RACG_DICDI STANDARD; PRT; 201 AA.
 AC 09GSP0:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE RAS-related protein racg.
 GN RACG.
 OS Dictyostellium discoideum (Slime mold).
 CC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
 OX NCBI_Taxid=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=21127961; PubMed=11222756;
 RA Rivero F., Diallich H., Glockner G., Noegel A.A.;
 RL "The Dictyostellium discoideum family of Rho-related proteins.";
 RT Nucleic Acids Res. 29:1068-1079(2001).
 CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF310893; AAC45130.1; -
 DR HSSP; P2181; 1AM4.
 DR Dictydb; DD27272; racg.
 DR InterPro; IPR003578; GTPase_Rho.
 DR InterPro; IPR001806; Ras_Transfmr.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00071; Ras; 1.
 DR PRINTS; PRO0449; RASTRNSFRNG.
 DR SMART; SM00174; RHO; 1.
 DR TIGRfam; TIGR00231; Email_GTP; 1.
 KW GTP-binding; Prenylation; Lipoprotein.
 FT NP_BIND 10 17 GTP (BY SIMILARITY).
 FT NP_BIND 57 61 GTP (BY SIMILARITY).
 FT NP_BIND 115 118 GTP (BY SIMILARITY).
 FT DOMAIN 32 40 EFFECTOR REGION (POTENTIAL).
 FT LIPID 198 198 GERANYL-GERANYL (BY SIMILARITY).
 SQ SEQUENCE 201 AA; 22613 MW; F6CE586497CC169 CRC64;

Query Match 5.6%; Score 75; DB 1; Length 201;
 Best Local Similarity 22.6%; Pred. No. 40;

Matches 36; Conservative 20; Mismatches 59; Indels 44; Gaps 7;

OY 44 MSFVVLGCVLSRDAQOEQLGRHYSQNDLV-EYSP-VTEKHLLDGMVRLCSLA 101
 DB 1 MRSIVCVVG-----EGGIGKTSMLSTYSNSISNEYQPTVFQNYST----- 42
 OY 102 ITMSDNTANLLTTIGGPKELTAFLHNMGDVTRLDREPELLEAIPNDEROTTTPYAMAT 145
 DB 43 LLMHKKPYNLSIMDTAGDEEFSKRLRLSTPQDVFLLCSLNPSSRNILDSVOELN 102
 OY 146 EAIPNDERDTPPVANATTLRLKLTGEL--TLASROOL 182
 DB 103 ENCPR-----TPVLGTQMDLKSNSVILDRLECKNQL 135

RESULT 3
 NDK_TREPA
 ID NDK_TREPA STANDARD; PRT; 149 AA.
 AC 08374:
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
 DE (Nucleoside-2-P kinase).
 GN NDK OR TP1010.
 OS Treponema pallidum.
 CC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_Taxid=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterback T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete.";
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
 CC OTHER THAN ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +
 CC nucleoside triphosphate.
 CC -1- SUBUNIT: Homotrimer (by similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- SIMILARITY: Belongs to the NDK family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE001269; AAC65961.1; -
 DR PIR; F71252; F71252.
 DR HSSP; P22887; INPK.
 DR TIGR; TP1010; -
 DR HAMAP; MF_00451; -; 1.
 DR InterPro; IPR001564; NDK.
 DR Pfam; PF00334; NDK; 1.
 DR PRINTS; PRO1243; NUCCDPKINASE.
 DR PRODOM; PD001018; NDK; 1.
 DR SMART; SM00562; NDK; 1.
 DR PROSITE; PS00469; NDP_KINASES; 1.
 KW Transferase; Kinase; ATP-binding; Complete proteome.
 FT ACT_SITE 117 117 BY SIMILARITY.
 SQ SEQUENCE 149 AA; 16657 MW; BB329539671E80BE CRC64;

```

RESULT 4
RAC2_LOTJUA
ID RAC2_LOTJUA STANDARD; PRT; 196 AA.
AC 040220;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RAC2-like GTP binding protein RAC2.
GN
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
OX NCBI_TaxID=34305;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Gifu / B-129; TISSUE=Root nodules;
RA MEDLINE=97231679; PubMed=9076991;
RX Bord S., Brandstrup B., Jensen T.J., Poulsen C.;
RT Identification of new protein species among 33 different small GTP-
RT binding proteins encoded by cDNAs from Lotus japonicus, and
RT expression of corresponding mRNAs in developing root nodules.*;
RL Plant J. 11:337-350(1997).
CC
CC -1- SIMILARITY: BELONGS TO THE SMALL GTPASE SUPERFAMILY. RHO FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as the content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

Db      1  MSTANFIK-----VTYGGAGANGKCTCLISYTSNTPFDIYP-----TFVDFNFS 44
Qy      100  AAIWMSDNTANALLTTTGGPKK-----LTAFL-LHNMGDHYRLDRREPE 144
Db      45  ANVVAV-DSGVTVNLGLWDVAGQEDVNRPLRPLSRGADVFLNARSLLSRASVENSKRIPE 100
Qy      144  LNEAIPIPNDRDTTPPAAMATTLKLLTGELLTL-ASRQOLLDMWEADKVA---GPLLRSA 195
Db      104  LRHYAP-----TVPF-----VLVGTKLIDREDRQYLIDHPGATPPTTAQGEELKKA 144
Qy      200  LPAGWFI 206
Db      150  IGAAYVL 156

RESULT 5
HC3L_THIFE
ID      HC3L_THIFE      STANDARD;      PRT;      122 AA.
AC      P80509;
DT      01-OCT-1996 (Rel. 34, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      Cytochrome-c3 hydrgenase, large chain (Ec 1.12.2.1) (Hydrogenase)
DE      (Ferments) .
GN      HONG.
OS      Thiobacillus ferrooxidans.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
OC      Acidithiobacillaceae; Acidithiobacillus.
OX      NCBI_Taxid:920;
RN      [1]
RP      SEQUENCE.
RC      STRAIN-AVCC 19859;
RX      MEDLINE-96241862; PubMed-8661919;
RA      Fischer J., Quentmeyer A., Kosika S., Kraft R., Friedrich C.G.;
RT      "Purification and characterization of the hydrogenase from
RL      Thiobacillus ferrooxidans."
RL      Arch. Microbiol. 165:289-296(1996).
CC      -1- CARBOLYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 - 2 H(+) + 2
CC      ferrocyclochrome c3.
CC      -1- COFACTOR: IRON.
CC      -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
CC      ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
KW      Oxidoreductase; Iron.
FT      NON_CONS      20      21
FT      NON_CONS      29      30
FT      NON_CONS      35      36
FT      NON_CONS      42      43
FT      NON_CONS      59      60
FT      NON_CONS      72      73
FT      NON_CONS      76      79
FT      NON_CONS      87      88
FT      NON_CONS      98      99
FT      NON_CONS      107     108
FT      NON_TER      122     122
SQ      SEQUENCE      122 AA; 13146 MW; EB4FAF365348DA0B CRC64;

Query Match      5.0%; Score 67; DB 1; Length 122;
Best Local Similarity 26.4%; Pred. No. 93;
Matches 28; Conservative 13; Mismatches 23; Indels 42; Gaps 4
Qy      19  VGIYELDNGSELSESRPERPMMSTFFKVLCAVYLSRIDAGQQLGRIRHYSQNDLV 78
Db      12  VGRVSGDLDSSKSLLEYFR-----NAILARFGGG-----LG 41
Qy      79  EYSPTEKHLTDGMTVRELCSAATMSDNTANALLTTTGGKRELT 124
Db      42  KYAPFTGTNYEIGVT-----ISGDKDPQAGLVVT---PREST 75

RESULT 6
RBFAL_UREPA
ID      RBFAL_UREPA      STANDARD;      PRT;      116 AA.
AC      Q9P0H0.

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR U0321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria: Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN SEQUENCE FROM N.A.
RP STRAIN=serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Casella G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE002129; AAF30730.1;
CC HAMAP: MF_000037; 1.
CC InterPro: IPR000238; Rib_bind_factA.
CC Pfam: PF02033; RBFA. 1.
CC ProDom: PD007327; Rib_bind_factA; 1.
CC TRIGRAMS: TIGR00082; rbfA; 1.
CC PROSITE: PS01319; RBFA; FALSE_NEG.
CC RNA processing; Complete proteome.
CC KW SEQUENCE 116 AA; 13247 MW; A79DCCT1F0547514 CRC64;
SQ
Query Match 4.9%; Score 66.5; DB 1; Length 116;
Best Local Similarity 28.9%; Pred. No. 96;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;
QY 144 LNEAIPNDERDTTPVAMATLKLITGEL-----LTLASROOLIDWME-ADKVAQPL 196
DB 18 INMLANEINDKIARLAVTAVR--LSNDLSVAKIFLDAHKRESMLKYLENVNRSVSG-LT 74
QY 197 RSLAPAGW-----FIADKS 210
DB 75 RSKLAEMWTSYKVPRLRFVIDET 97

```

```

RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerevage A.R., Graham D.E., Kyriides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner S.E., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Welch C.I., McNeil L.R., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
CC guanine (7,8-dihydro-8-oxoguanine = 7-oxoG) from DNA. Also nicks
CC DNA at apurinic/aprimidinic sites (AP sites) (by similarity).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE001079; AAB9876.1;
CC PIR: C69296; C69296.
CC TIGR: AF0371;
CC HAMAP: MF_00241; 1.
CC InterPro: IPR003265; Endo_3c.
CC Pfam: PF00730; Hnh-GPD; 1.
CC SMART: SM00478; ENDO3c; 1.
CC DR Hydrolyase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
CC KW Multifunctional enzyme; Complete proteome.
CC FT ACT_SITE 122
CC SEQUENCE 198 AA; 22639 MW; 3A5C033AA12F3FFB CRC64;
SQ
Query Match 4.9%; Score 65.5; DB 1; Length 198;
Best Local Similarity 25.0%; Pred. No. 2,3e+02;
Matches 36; Conservative 19; Mismatches 54; Indels 35; Gaps 7;
QY 14 QLCARNGYIELDNLNGELIESFREPERPMSTFVLLGAVL---SRIDAG---OEOG 67
DB 15 QLCERK-----GEVEFDFRPFLLDFSVATITTELAFCISTANSSATGAKFORLLG 64
QY 68 RRIHYDQ-----NDIVEYSPVTEKHLTDGATVELCSAAITMSDNTANLLTTI 117
DB 65 QGVGVKRALTLGAYRPHNRKAEY-----IEAKRFRVLKALAEASKRAEILLIKR 117
QY 118 G-GPKETLAPLHNMG-DHYTRIDR 139
DB 118 GLGKKEASHPLRNVGRDVAIIDR 141

```

```

RESULT 7
OGGI_ARCFU
ID OGG1_ARCFU STANDARD; PRT; 198 AA.
AC 029876;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable N-glycosylase/DNA lyase [includes: 8-oxoguanine DNA
DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
DE (EC 4.2.99.18) (AP lyase)].
GN OGG OR AF0371.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN SEQUENCE FROM N.A.
RP STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC

```

```

RESULT 8
KTHY_METAC
ID KTHY_METAC STANDARD; PRT; 206 AA.
AC 08785;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
GN TMK OR M4433.
OS Methanosarcina acetivorans.
OC Archaea: Euryarchaeota; Euryarchaeota orders Incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.

```


OX NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., MacDonald P.,
 RA Pitzhugh W., Calvo S., Engels R., Smirnov S., Atencio D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Lincon L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grathme D.A., Guss A.M.,
 RA Hederich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanocaldococcus jassbyi reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
 CC 5'-diphosphate.
 CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE011164; AM07774.1; -
 DR HAMAP: MF_00165; -1.
 DR InterPro: IPR000062; Thymidylate_kin.
 DR Pfam: PF02223; Thymidylate_kin; 1.
 DR TIGRFAMs: TIGR00041; DIMP_kinase; 1.
 DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSR_NRG.
 DR Trasnferase; Kinase; Nucleotide biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP_BIND
 FT BIND 10 17 ATP (POTENTIAL).
 SQ SEQUENCE 206 AA; 23303 MW; A19CE5BC85423B3 CRC64;
 Query Match 4.9%; Score 65.5; DB 1; Length 206;
 Best Local Similarity 18.7%; Pred. No. 2.4e+02;
 Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;
 QY 56 LSRID-AGOEOLGRIRHYSQNDLVEYSPYTEKHLTDGTVNELCSAITSMDNTAAILL 114
 DB 8 LEGIDGSGKSTFAKKIQ-ENSELRVPEVPTREPTRG-----TLGNNAVENAIQ 55
 QY 115 TTIGSPKELTAFVLMHMGDVTNLRDREPELNEAIPNDEKOTTTTPVAMATTIRKILTGEL 174
 DB 56 SDTDLALFLFTADHAELAKLVKPALEGGKTVISDRYSRSRYAOGITLKNRLDNPJ- 114
 QY 175 TIASROQLIDWMEADKVAGPLLRSLAPAGW-----FIAD-----KSG-AGE----- 214
 DB 115 -----EMVR-----DLHKGTVIPLDLTFLDIEPEIAVKRCGKRGGEQTKF 154
 QY 215 -----KSGRTIALAGDPGKPSRIYVITTSQATNDE 247
 DB 155 EKIEFLNGVRELFLGLAAE-EPERFYIVDASGSPDEYER 192
 RESULT 9
 VNSC_RINDK STANDARD; PRT; 177 AA.
 AC P35948;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Nonstructural protein C.
 GN C.
 OS Rinderpest virus (strain Kabete 0) (RDV).

CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
 CC NCBI_TaxID=11242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92410646; PubMed=1529555;
 RA Yamanaoka M., Dale B., Crisp T., Cordell B., Grubman M., Ylma T.;
 RT "Sequence analysis and editing of the phosphoprotein (P) gene of
 RT rinderpest virus.";
 RL Virology 190:553-556(1992).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S44819; AAB3269.1; -
 DR PIR: B43387; B43387.
 DR InterPro: IPR003875; Paramyxovir_NSC.
 DR Pfam: PF02725; Paramyxo_NSC; 1.
 KW Nonstructural protein.
 SQ SEQUENCE 177 AA; 19926 MW; 116971B140A39F11 CRC64;
 Query Match 4.8%; Score 65; DB 1; Length 177;
 Best Local Similarity 24.6%; Pred. No. 2.2e+02;
 Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;
 QY 93 TVRELCSAITSMDNTAAILLTITIGSPKELTAFVLMHMGDVTNLRDREPELNEAIPNDE 152
 DB 49 TIRISASHSNQQLDQKACIAVTIRDLKATVMSR-----WEHL----- 90
 QY 153 RDTTP-----VAATTIRKILTGELTLASQQLIDWMEADKVAGPLLRSLAP 202
 DB 91 ---VPOCTAPRYSITIMFTAVKRLRESKMLTISFNQAL--MNVSR-SGSEKRNILRTA 144
 QY 203 GWFIND 208
 DB 145 MILAN 150
 RESULT 10
 VNSC_RINDR STANDARD; PRT; 177 AA.
 AC Q03339;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Nonstructural protein C.
 GN C.
 OS Rinderpest virus (strain RHOK) (RDV).
 CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
 CC NCBI_TaxID=36409;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93155658; PubMed=8429304;
 RA Baron M.D., Shalla M.S., Barrett T.;
 RT "Cloning and sequence analysis of the phosphoprotein gene of
 RT rinderpest virus.";
 RL J. Gen. Virol. 74:299-304(1993).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68311; CAA48391.1; -


```

OC Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxId=2336;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.C., Dodson R.J.,
RA Hart D.H., Hickey E.K., Peterson J.D., Nelson W.L., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.H.,
RA Stear A.M., Sutton M.D., Pratt M.S., Phillips G.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Protease subunit of a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-slb.ch/announce/
CC or send an email to license@lsb-slb.ch).
CC
CC CC EMBL, AE001728; AACD5606.1; -.
CC DR PIR: G72365; G72365.
CC DR HSSP: P31059; 1E94.
CC DR MEROPS: T01.006; -.
CC DR TIGR: TM0521; -.
CC DR HAMAP: MF_00248; -; 1.
CC DR InterPro: IPR001353; Protsme.protease.
CC DR Pfam: PF00227; proteasome; 1.
CC KW Hydrolase; Protease; Complete proteome.
CC FT ACT_SITE 6
CC FT SITE 6 BY SIMILARITY.
CC SQ SEQUENCE 176 AA; 18933 MW; EC6369602ADABD02 CRC64;

```

```

RP  SEQUENCE FROM N.A.
RC  STRAIN-LT2;
RX  MEDLINE-99429843; PubMed-10498708;
RA  Bobik T.A., Havemann G.D., Busch R.J., Williams D.S., Aldrich H.C.;
RT  "The propanediol utilization (pdu) operon of Salmonella enterica
RT  serovar typhimurium LT2 includes genes necessary for formation of
RT  polyphedral organelles involved in coenzyme B(12)-dependent 1, 2-
RT  propanediol degradation."
RL  J. Bacteriol. 181:5967-5975(1999).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN-LT2 / SGC1412 / ATCC 700720;
RX  MEDLINE-21534948; PubMed-11677609;
RA  McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA  Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA  Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA  Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA  Waterston R., Wilson R.K.;
RT  "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT  LT2."
RL  Nature 413:852-856(2001).
CC  -1- INDUCTION: BY propanediol.
CC  -1- SIMILARITY: BELONGS TO THE EUTP/PDUV FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: AF026270; AAD39020.1; -
DR  EMBL: AE008790; AAL20960.1; -.
DR  StGene: SGI0580; pduV.
KM  ATP-binding; Complete proteome.
FT  NP_BIND 8 15
SO  SEQUENCE 150 AA; 16348 MW; 7771229432f97e56 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 150;
Best Local Similarity 19.2%; Pred. No. 2.8e+02;
Matches 29; Conservative 20; Mismatches 75; Indels 27; Gaps 4;

QY  52 CGAVLSRIDAGQEQIGRIRHYSNDLYEVPYTEKHLTDGNTVRELCSAATMSDNTAAN 111
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  12 CG-----KTSIQSLSGEGALHFKTKQAIEWSPMAID--TPGEYLENRCIYSALITSACEAD 65
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  112 LLTTTGGPKELTAFLHNMGDHYTRLDRPEPELNEAIPNDEROTTTPVAMATTLKLTG 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db  66 VIALVLNADAGQMSPF-----SPGFTAPANRPITIGLYTKADLAEPORISLVA 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY  172 ELTLASRQQLIDRMWADRVAGPLRSALPA 202
   || | | | | | | | | | | | | | | | | | | | | | |
Db  112 EMLTQAGQGIIF-----ITSALNNSGLDA 135
   || | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: September 10, 2003, 12:20:51
 Job time : 5.57143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 25.4286 Seconds
(without alignments)
2668.960 Million cell updates/sec

Title: SEQ2_30E_157T
Perfect score: 1348
Sequence: 1 HPEELVKVKAEDQIGARVC.....TWDERNRQIAEIGASLIKHW 263

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 365421

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

SPREMBL_23:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 38.9 | 105 | 2 | 052026 |
| 2 | 524 | 38.9 | 105 | 2 | 052330 |
| 3 | 499 | 37.0 | 145 | 2 | 08RTD8 |
| 4 | 488 | 36.2 | 102 | 2 | 052639 |
| 5 | 464 | 34.4 | 138 | 2 | 0990F3 |
| 6 | 464 | 34.4 | 138 | 2 | 09AM1 |
| 7 | 464 | 34.4 | 138 | 2 | 09AM2 |
| 8 | 462 | 34.3 | 138 | 2 | 09AM0 |
| 9 | 462 | 34.3 | 139 | 2 | 09AM9 |
| 10 | 459 | 34.1 | 139 | 2 | 09AM3 |
| 11 | 459 | 34.1 | 139 | 2 | 09AM3 |
| 12 | 338 | 25.1 | 67 | 2 | 053553 |
| 13 | 326 | 24.2 | 95 | 2 | 086DES |
| 14 | 310 | 23.0 | 62 | 2 | 09JN58 |
| 15 | 301 | 22.3 | 128 | 2 | 08VQ00 |
| 16 | 251 | 18.6 | 180 | 2 | 08KVT2 |

| | | | | | | |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 246 | 18.2 | 52 | 2 | 09R412 | 09R412 shigella fl |
| 18 | 241 | 17.9 | 48 | 2 | 09RLH0 | 09RLH0 proteus mfr |
| 19 | 197 | 14.6 | 38 | 2 | 097145 | 097145 escherichia |
| 20 | 164.5 | 12.2 | 134 | 2 | 08VUL3 | 08VUL3 staphylococ |
| 21 | 119.5 | 8.9 | 109 | 2 | 053698 | 053698 staphylococ |
| 22 | 105 | 7.8 | 20 | 2 | 097146 | 097146 escherichia |
| 23 | 87 | 6.1 | 100 | 2 | 093S05 | 093S05 staphylococ |
| 24 | 82 | 6.1 | 68 | 2 | 09XB12 | 09XB12 bacillus ce |
| 25 | 80 | 5.9 | 198 | 2 | 09ACM8 | 09ACM8 streptococ |
| 26 | 79.5 | 5.9 | 202 | 16 | 0985Z0 | 0985Z0 trypsin |
| 27 | 75.5 | 5.6 | 190 | 5 | 076227 | 076227 trypsin |
| 28 | 75.5 | 5.6 | 200 | 16 | 097JH8 | 097JH8 clostridium |
| 29 | 75 | 5.6 | 181 | 16 | 08D108 | 08D108 yersinia pe |
| 30 | 74.5 | 5.5 | 113 | 2 | 049970 | 049970 mycobacteri |
| 31 | 74.5 | 5.5 | 172 | 16 | 08TBM3 | 08TBM3 brucella me |
| 32 | 74.5 | 5.5 | 172 | 16 | 08TBM3 | 08TBM3 brucella su |
| 33 | 73.5 | 5.5 | 195 | 16 | 09RIA0 | 09RIA0 streptococ |
| 34 | 73.5 | 5.5 | 204 | 11 | 091YW4 | 091YW4 mus musculu |
| 35 | 73 | 5.4 | 131 | 2 | 09X9H0 | 09X9H0 yersinia en |
| 36 | 73 | 5.4 | 153 | 5 | 08SSZ0 | 08SSZ0 clona lites |
| 37 | 73 | 5.4 | 175 | 17 | 08ZTH8 | 08ZTH8 pyrobaculum |
| 38 | 72.5 | 5.4 | 145 | 2 | 005984 | 005984 staphylococ |
| 39 | 72 | 5.3 | 205 | 16 | 09BJT8 | 09BJT8 rhizobium i |
| 40 | 71.5 | 5.3 | 152 | 16 | 09RI91 | 09RI91 streptococ |
| 41 | 71.5 | 5.3 | 196 | 2 | 09AH34 | 09AH34 pseudomonas |
| 42 | 71 | 5.3 | 150 | 10 | 0941G4 | 0941G4 nicotiana t |
| 43 | 71 | 5.3 | 153 | 16 | 09PKT4 | 09PKT4 chlamydia m |
| 44 | 71 | 5.3 | 177 | 16 | 09KEM7 | 09KEM7 bacillus ha |
| 45 | 71 | 5.3 | 192 | 16 | 09HW07 | 09HW07 pseudomonas |

ALIGNMENTS

RESULT 1

| | | | | |
|----|---|--------------|-----------|-------------------------|
| ID | 052026 | PRELIMINARY; | PRT; | 105 AA. |
| AC | 052026; | | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Created) | | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Last sequence update) | | | |
| DT | 01-MAR-2002 (TREMBLrel. 20, Last annotation update) | | | |
| DE | Beta-lactamase (Fragment). | | | |
| GN | BLA. | | | |
| OS | Pseudomonas aeruginosa. | | | |
| OG | Plasmid PR01614. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; | | | |
| OC | Pseudomonadaceae; Pseudomonas. | | | |
| OX | NCBI_TaxID=287; | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RX | MEDLINE=95011664; PubMed=7926843; | | | |
| RA | West S.E., Schweizer H.P., Dall C., Sample A.K., Runyen-Janecky L.J.; | | | |
| RT | "Construction of Improved Escherichia-Pseudomonas Shuttle Vectors | | | |
| RT | derived from pUC18/19 and sequence of the region required for their | | | |
| RT | replication in Pseudomonas aeruginosa."; | | | |
| RL | Gene 148:81-86(1994) | | | |
| DR | EMBL; L30112; AAA6058.1; - | | | |
| DR | HSSP; P00810; 1XPB. | | | |
| DR | InterPro; IPR001466; Beta_lactamase. | | | |
| DR | InterPro; IPR000871; Beta_lactamase_A. | | | |
| DR | Pfam; PF00144; beta-lactamase; 1. | | | |
| DR | PRINTS; PR00118; BLACTAMASEA. | | | |
| KW | Plasmid. | | | |
| FT | NON_TER | | | |
| SQ | SEQUENCE | 105 AA; | 11229 MW; | D2889AA073330557 CRC64; |

Query Match 38.9%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 1e-36;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTIRKLNGELLTLASROOIIDMMEADKVGPIRLSALPAGFTADKSGAGSGSG 219
Db 2 AMATTIRKLNGELLTLASROOIIDMMEADKVGPIRLSALPAGFTADKSGAGSGSG 61

QY 220 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 263
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 62 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105

RESULT 2
Q52330 PRELIMINARY; PRT; 105 AA.
AC Q52330.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Tnl bla protein (Fragment).
OS Escherichia coli.
OC Plasmid RK2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90264294; PubMed=2160936;
RA Kornacki J.A., Burlage R.S., Figurski D.H.;
RT "The kll-kor regulon of broad host-range Plasmid RK2: Nucleotide
RT sequence, polypeptide product and expression of regulatory gene
RT korC".
RL J. Bacteriol. 172:3040-3050(1990).
DR EMBL; M32794; AAA26408.1; .
DR HSSP; P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam; PF00144; beta_lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER
SQ SEQUENCE 105 AA; 11229 MW; D2869A407330557 CRC64;

Query Match 38.9%; Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 1e-36;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATLRKLTGELTLASROQLDWMKADKVPGLRSALPAGWFIADKSGAGERSRG 219
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 2 AMATLRKLTGELTLASROQLDWMKADKVPGLRSALPAGWFIADKSGAGERSRG 61

QY 220 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 263
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 62 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105

RESULT 3
Q52330 PRELIMINARY; PRT; 145 AA.
AC Q52330.
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE SHV-5 enzyme (Fragment).
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX Palasubdranlam S.;
RT "Characterization of SHV-5 extended-spectrum beta-lactamase from
RT ceftazidime-resistant Klebsiella pneumoniae".
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF467105; AAL75506.1; .
DR InterPro: IPR001466; Beta_lactamase.
DR Pfam; PF00144; beta_lactamase; 1.
FT NON_TER
SQ SEQUENCE 145 AA; 145 MW; D2869A407330557 CRC64;

SQ SEQUENCE 145 AA; 15574 MW; F88634D6194B4C82 CRC64;

Query Match 37.0%; Score 499; DB 2; Length 145;
Best Local Similarity 68.8%; Pred. No. 2.1e-34;
Matches 95; Conservative 18; Mismatches 25; Indels 0; Gaps 0;

QY 100 AAITMSDNTAANLLTFTIGGKELTAFTHNMGDIYTRDRWPELNEAIPNDRPTTV 159
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 7 AAITMSDNTAANLLTFTIGGKELTAFTHNMGDIYTRDRWPELNEAIPNDRPTTV 66

QY 160 AMATLRKLTGELTLASROQLDWMKADKVPGLRSALPAGWFIADKSGAGERSRG 219
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 67 SMAATLRKLTGELTLASROQLDWMKADKVPGLRSALPAGWFIADKSGAGERSRG 126

QY 220 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 263
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 127 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105

RESULT 4
Q52639 PRELIMINARY; PRT; 102 AA.
AC Q52639.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Beta-lactamase (Fragment).
GN BLA.
OS Pseudomonas aeruginosa.
OC Plasmid pRO1600.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX TRANSPOSON-Tn1;
RC MEDLINE=94336757; PubMed=8058819;
RA Jansons I., Touchle G., Sharp R., Almquist K., Farinha M.A., Lam J.S.,
RA Kropinski A.M.;
RT "Deletion and transposon mutagenesis and sequence analysis of the
RT pRO1600 oriR region found in the broad-host-range plasmids of the pOF
RT series".
RL Plasmid 31:265-274(1994).
DR EMBL; L22691; AAA98312.1; .
DR HSSP; P00810; 1XPB.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam; PF00144; beta_lactamase; 1.
DR PRINTS; PR00118; BLACTAMASEA.
KW Plasmid.
FT NON_TER
SQ SEQUENCE 102 AA; 10827 MW; DA95AF7557DA1304 CRC64;

Query Match 36.2%; Score 488; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.1e-33;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 AMATLRKLTGELTLASROQLDWMKADKVPGLRSALPAGWFIADKSGAGERSRG 219
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 2 AMATLRKLTGELTLASROQLDWMKADKVPGLRSALPAGWFIADKSGAGERSRG 61

QY 220 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 263
IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105
Db 62 IIAALGPDGKPSRIVYITTTGSAATMDERNROIAEIGASLTKHW 105

RESULT 5
Q52639 PRELIMINARY; PRT; 138 AA.
AC Q52639.
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-6988, and E/99-5-2;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327049; AK07464.1; -
DR EMBL; AF327051; AK07466.1; -
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase_1.
DR PRINTS; PR00118; BLACTMASEA.
FT NON_TER 1 1
SQ SEQUENCE 138 AA; 15105 MW; 561D09282442A847 CRC64;

Query Match 34.4%; Score 464; DB 2; Length 138;
Best local Similarity 68.5%; Pred. No. 1.8e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLITIGGPKELTAFILNMGDHYTRLDKWEPELNEAIPNDERDTTPVAMATTLRLKLT 170
DB 10 NLLATVGGPAGLTFALRIGDNYTRLDKWEPELNEALPGDARDTTPASMAATLRLKLT 69
OY 171 GELLTASROOLDIMMEADKVAGPLLRSLPAGWFTADKSGAGRGSGITIAALGPDKP 230
DB 70 SQRLSARSOROLDMMVDDRVAGPLIRSVLPAGWFTADKSGARGIVALLGPNNKA 129
OY 231 SRIYVY 237
DB 130 ERIYVY 136

RESULT 6

O9AMA1 PRELIMINARY; PRT; 138 AA.

AC O9AMA1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-18/25;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327050; AK07465.1; -
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase_1.
DR PRINTS; PR00118; BLACTMASEA.
FT NON_TER 1 1
SQ SEQUENCE 138 AA; 15204 MW; 56094C3C7507BC02 CRC64;

Query Match 34.4%; Score 464; DB 2; Length 138;
Best local Similarity 68.5%; Pred. No. 1.8e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLITIGGPKELTAFILNMGDHYTRLDKWEPELNEAIPNDERDTTPVAMATTLRLKLT 170
DB 10 NLLATVGGPAGLTFALRIGDNYTRLDKWEPELNEALPGDARDTTPASMAATLRLKLT 69
OY 171 GELLTASROOLDIMMEADKVAGPLLRSLPAGWFTADKSGAGRGSGITIAALGPDKP 230
DB 70 SQRLSARSOROLDMMVDDRVAGPLIRSVLPAGWFTADKSGARGIVALLGPNNKA 129
OY 231 SRIYVY 237
DB 130 ERIYVY 136

RESULT 7

O9AMA2 PRELIMINARY; PRT; 139 AA.

AC O9AMA2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-918;
RA Subramaniam G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant Escherichia coli
isolates from UMC, Malaysia.*;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF327048; AK07463.1; -
DR HSSP; P14557; ISHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase_1.
DR PRINTS; PR00118; BLACTMASEA.
FT NON_TER 1 1
SQ SEQUENCE 139 AA; 15234 MW; 0B161D09282442A8 CRC64;

Query Match 34.4%; Score 464; DB 2; Length 139;
Best local Similarity 68.5%; Pred. No. 1.8e-31;
Matches 87; Conservative 17; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLITIGGPKELTAFILNMGDHYTRLDKWEPELNEAIPNDERDTTPVAMATTLRLKLT 170
DB 10 NLLATVGGPAGLTFALRIGDNYTRLDKWEPELNEALPGDARDTTPASMAATLRLKLT 69
OY 171 GELLTASROOLDIMMEADKVAGPLLRSLPAGWFTADKSGAGRGSGITIAALGPDKP 230
DB 70 SQRLSARSOROLDMMVDDRVAGPLIRSVLPAGWFTADKSGARGIVALLGPNNKA 129
OY 231 SRIYVY 237
DB 130 ERIYVY 136

RESULT 8

O9AMA0 PRELIMINARY; PRT; 138 AA.

AC O9AMA0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-E/98 9-1;
RA Subramanian G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant *Escherichia coli*
RT Isolates from UMMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327052; AAK07467.1; -
DR HSSP: P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 138 138
FT SEQUENCE 138 AA; 15105 MW; 561D092F5442A847 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 2.6e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLHNGDHTRLDRWEPELNEALPNDERDTTPVAMATTLKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTTPASMAATLKLIT 69
OY 171 GELLTLASRQQLIDMWEADRVAGPLRSALPGWFIADKSGAGERSGRIITAAIGPDGR 230
DB 70 SQRLSARSGROLQMWVDVAVAGPLRSVLPAGWFIADKTSKRGARGIVALLGPNMKA 129
OY 231 SRIVVY 237
DB 130 ERIVVLY 136

RESULT 9
O9AM99 PRELIMINARY; PRT; 138 AA.

AC O9AM99;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/99 4-1;
RA Subramanian G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant *Escherichia coli*
RT Isolates from UMMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327053; AAK07468.1; -
DR HSSP: P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 138 138
FT SEQUENCE 138 AA; 15204 MW; 56094C3B0507BC02 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 138;
Best Local Similarity 67.7%; Pred. No. 2.6e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLHNGDHTRLDRWEPELNEALPNDERDTTPVAMATTLKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTTPASMAATLKLIT 69
OY 171 GELLTLASRQQLIDMWEADRVAGPLRSALPGWFIADKSGAGERSGRIITAAIGPDGR 230
DB 70 SQRLSARSGROLQMWVDVAVAGPLRSVLPAGWFIADKTSKRGARGIVALLGPNMKA 129

OY 231 SRIVVY 237
DB 130 ERIVVLY 136

RESULT 10

O9AM98 PRELIMINARY; PRT; 139 AA.

AC O9AM98;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E/99 3-2;
RA Subramanian G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant *Escherichia coli*
RT Isolates from UMMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327054; AAK07469.1; -
DR HSSP: P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.
DR InterPro: IPR000871; Beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00118; BLACTAMASEA.
FT NON_TER 139 139
FT SEQUENCE 139 AA; 15234 MW; 0361A792F5442A8 CRC64;

Query Match 34.3%; Score 462; DB 2; Length 139;
Best Local Similarity 67.7%; Pred. No. 2.6e-31;
Matches 86; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

OY 111 NLLTTTGGPKELTAFLHNGDHTRLDRWEPELNEALPNDERDTTPVAMATTLKLLT 170
DB 10 NLLLATVGGPAGLTAFLRQIGDVTRLDRWETELNEALPGARDTTTPASMAATLKLIT 69
OY 171 GELLTLASRQQLIDMWEADRVAGPLRSALPGWFIADKSGAGERSGRIITAAIGPDGR 230
DB 70 SQRLSARSGROLQMWVDVAVAGPLRSVLPAGWFIADKTSKRGARGIVALLGPNMKA 129
OY 231 SRIVVY 237
DB 130 ERIVVLY 136

RESULT 11

O9AM93 PRELIMINARY; PRT; 139 AA.

AC O9AM93;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE SHV beta-lactamase (Fragment).
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-935;
RA Subramanian G., Navaratnam P.;
RT *SHV gene sequences from ceftazidime-resistant *Escherichia coli*
RT Isolates from UMMC, Malaysia.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327047; AAK07462.1; -
DR HSSP: P14557; 1SHV.
DR InterPro: IPR001466; Beta_lactamase.

DR InterPro: IPR000871; beta_lactamase_A.
DR Pfam: PF00144; beta-lactamase; 1.
DR PRINTS: PR00116; BLACTMASEA.
FT NON_TER 1
FT NON_TER 139
SQ SEQUENCE 139 AA; 15257 MW; E514247C882442AD CRC64;

Db 61 AG 62

RESULT 15

000000

| | | | | |
|----|--------|--------------|------|---------|
| ID | Q8VQ00 | PRELIMINARY; | PRT; | 128 AA. |
| | Q8VQ00 | | | |

DT 01-MAR-2002 (Tremblay, 20, Created)

| | | |
|----|-------------|-------------------------------------|
| DT | 01-MAR-2002 | (TREMBlé, 20, Last sequence update) |
| DT | 01-MAR-2003 | (TREMBlé, 23, Last sequence update) |

01-MAR-2003 (EMBL/EBL. 23, Last annotation update,
CTX-M type beta-lactamase (Fragment))

Escherichia coli.

Bacteria; Proteobacteria

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

| | |
|----|-----|
| RN | [1] |
|----|-----|

RP SEQUENCE FROM N.A.

| | |
|----|------------------------------|
| RC | STRAIN-clinical isolate HD3; |
| PA | Stuenkelburg E Feucht H I |

A new CTX-M type beta-lactamase:

submitted (JAN-2002) to the EMBL/Gen

DR EMBL; AF466148; AAL73983.1; -.

DR InterPro; IPR001466; Beta_lactamase

DR InterPro; IPR000871; beta_lactamase

DR Pfam; PF00144; beta-lactamase; 1.

DR PRINTS; PRO0118; BLACTAMASEA.
ET NON TED 1

| | | |
|---------|---|-----|
| ET | 1 | 128 |
| NON_TER | 1 | 128 |
| NON_TER | 1 | 128 |
| ET | 1 | 128 |

| NO. | NAME | AGE |
|-----|------|-----|
| 1 | JOHN | 12 |
| 2 | JANE | 10 |
| 3 | JOHN | 12 |
| 4 | JANE | 10 |
| 5 | JOHN | 12 |
| 6 | JANE | 10 |
| 7 | JOHN | 12 |
| 8 | JANE | 10 |
| 9 | JOHN | 12 |
| 10 | JANE | 10 |
| 11 | JOHN | 12 |
| 12 | JANE | 10 |
| 13 | JOHN | 12 |
| 14 | JANE | 10 |
| 15 | JOHN | 12 |
| 16 | JANE | 10 |
| 17 | JOHN | 12 |
| 18 | JANE | 10 |
| 19 | JOHN | 12 |
| 20 | JANE | 10 |
| 21 | JOHN | 12 |
| 22 | JANE | 10 |
| 23 | JOHN | 12 |
| 24 | JANE | 10 |
| 25 | JOHN | 12 |
| 26 | JANE | 10 |
| 27 | JOHN | 12 |
| 28 | JANE | 10 |
| 29 | JOHN | 12 |
| 30 | JANE | 10 |
| 31 | JOHN | 12 |
| 32 | JANE | 10 |
| 33 | JOHN | 12 |
| 34 | JANE | 10 |
| 35 | JOHN | 12 |
| 36 | JANE | 10 |
| 37 | JOHN | 12 |
| 38 | JANE | 10 |
| 39 | JOHN | 12 |
| 40 | JANE | 10 |
| 41 | JOHN | 12 |
| 42 | JANE | 10 |
| 43 | JOHN | 12 |
| 44 | JANE | 10 |
| 45 | JOHN | 12 |
| 46 | JANE | 10 |
| 47 | JOHN | 12 |
| 48 | JANE | 10 |
| 49 | JOHN | 12 |
| 50 | JANE | 10 |

Query match 22.38; Score 301; DB 2; Length 128;

Best Local Similarity 48.4%; Pred. NO. 8.9e-18;

Matches 59; Conservative 20; Mismatches 43; Indels 0; Gaps 0;

66 LGRIHXSQNDLVEYSPVTEKHLTDGNTVRELSAAITMSDNTAANLLTTIGGPKELTA 125

6 INRVETKSDIYNYNPATAEKHYNGTMSIAETSAATIOYSDNYAMNKI.TAHVGGPASYTEE

126 FI.HNMGDHTRI.DWEPEI.NEATPNDEDTTTPVAMATTI.PKLTGEJLTI.ASPROO.ITW 185

120 F LHMVCHV I ANLNMV F E DNEH I F NDEH I I F VLVH I I LNKH I I GJHH I LMSKVH I DHH 100

Db 66 FARQLE

Search completed: September 10, 2003, 12:29:29
Job time : 26.4286 secs

THIS PAGE BLANK (USPTO)

SO SEQUENCE 176 AA; 18933 MW; ECE369602A0AED02 CRC64;

Query Match 4.7%; Score 63; DB 1; Length 176;

Best Local Similarity 28.2%; Pred. No. 3e+02; Mismatches 35; Conservative 12; Mismatches 31; Indels 46; Gaps 7;

QY 117 IGGPKELTAFLEHNGDHYTRLDREPELNEAI PNDERDTTPVAMATIRKLTGLTTL 176
 DB 42 LGGKVIAGPAGSVAADMTLEDFREAKRE-----WGNLTK 78
 QY 177 ASGQQLIDMEAKVACPLRSALPACWFIADK-----SGAGGSGSGII-----AAL 224
 DB 79 AAVEALKDW-RTDRV-----LR-RLEALLVADKENIFITISNGE-----VIGPDDDAAI 127
 QY 225 GPDG 228
 DB 128 GSGG 131

RESULT 14

ID PDUV_SALTY STANDARD; PRT; 150 AA.

AC 09XDM6; 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Propanediol utilization protein pduv.
 GN PDUV OR STM2056.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RA MEDLINE=99429843; PubMed=10498708;
 RA Bobik T.A., Havemann G.D., Busch R.J., Williams D.S., Aldrich H.C.;
 RT "The propanediol utilization (pdu) operon of Salmonella enterica
 RT serovar typhimurium LT2 includes genes necessary for formation of
 RT polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-
 RT propanediol degradation.";
 RL J. Bacteriol. 181:5967-5975(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 CC -1- INDUCTION: BY propanediol.
 CC -1- SIMILARITY: BELONGS TO THE EUTP/PDUV FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF062270; AAD38020.1; -;
 DR EMBL: AE008790; AAL20960.1; -;
 DR StyGene: SG10680; pduv.
 KW ATP-binding; Complete proteome.
 FT NP_BIND 8
 SO SEQUENCE 150 AA; 16348 MW; 7771229432F97E56 CRC64;

Query Match 4.6%; Score 62.5; DB 1; Length 150;

Best Local Similarity 19.2%; Pred. No. 2.7e+02; Mismatches 29; Conservative 20; Mismatches 75; Indels 27; Gaps 4;

QY 52 CGAVLSRIDAGQPOLGRIRHSONDLVEYSPVTEKHLWDGMVRELCSAIIIMSDMTAN 111
 DB 12 CG-----KTSLSQSLRGALHKKTOAIEWSMAID--TPGEYLRNCLYSALTSCAAD 65
 QY 112 LITTTGGPKELTAFLEHNGDHYTRLDREPELNEAI PNDERDTTPVAMATIRKLTGL 171
 DB 66 VIALVYNADQNSPF-----SPGFAPMNRPTIGVTAKDLAPORISLVA 111
 QY 172 ELTTLASROQLIMMEADKVAGPLRSALPA 202
 DB 112 EMLTGAQOIF-----ITSALNSGLDA 135

RESULT 15

ID YDEJ_ECOLI STANDARD; PRT; 172 AA.

AC P31131; 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein ydeJ.
 GN YDEJ OR B1537.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=93186717; PubMed=8383113;
 RA Cohen S.P., Haechler H., Levy S.B.;
 RT "Genetic and functional analysis of the multiple antibiotic
 RT resistance (mar) locus in Escherichia coli.";
 RL J. Bacteriol. 175:1484-1492(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alpa H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sempel G., Seki Y., Sivasubraman S.,
 RA Tagami H., Takeuchi Y., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 CC -1- SIMILARITY: BELONGS TO THE CINA FAMILY. STRONG, TO E. COLI YGAD.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M96235; -; NOT ANNOTATED_CDS.
 DR EMBL: AE000251; AAC74610.1; -;
 DR EMBL: D90796; BAA15227.1; -;

| ID | YCJC_ECOLI | STANDARD: | PRT: | 185 AA. |
|----|--|-----------------------------------|---------|---------|
| AC | P38522: | P76839: | P77417: | |
| DT | 01-OCT-1994 | (Rel. 30, Created) | | |
| DT | 01-NOV-1997 | (Rel. 35, Last sequence update) | | |
| DT | 28-FEB-2003 | (Rel. 41, Last annotation update) | | |
| DE | Hypothetical protein YcJc. | | | |
| GN | YCJC OR B1299 OR SFI304. | | | |
| OS | Escherichia coli, and | | | |
| OS | Shigella flexneri. | | | |
| OC | Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; | | | |
| OC | Enterobacteriaceae; Escherichia. | | | |
| OX | NCBI_Taxid=562, 623; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | SPECIES=E.coli; STRAIN=K12 / MG1655; | | | |
| RX | MEDLINE=97426617; PubMed=9278503; | | | |
| RA | Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., | | | |
| RA | Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., | | | |
| RA | Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., | | | |
| RA | Mau B., Shao Y.; | | | |
| RT | "The complete genome sequence of Escherichia coli K-12."; | | | |
| RL | Science 277:1453-1474(1997). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | SPECIES=E.coli; STRAIN=K12; | | | |
| RX | MEDLINE=97251357; PubMed=9097039; | | | |
| RA | Alba H., Bada T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T., | | | |
| RA | Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M., | | | |
| RA | Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., | | | |
| RA | Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N., | | | |
| RA | Takemoto K., Takeuchi Y., Sivasubraman S., Tagami H., Takeda J., | | | |
| RA | Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.; | | | |
| RT | "A 570-kb DNA sequence of the Escherichia coli K-12 genome | | | |
| RT | corresponding to the 28.0-40.1 min region on the linkage map."; | | | |
| RL | DNA Res. 3:363-377(1996). | | | |
| RN | [3] | | | |
| RP | SEQUENCE OF 65-185 FROM N.A. | | | |
| RC | SPECIES=E.coli; | | | |
| RX | MEDLINE=91216440; PubMed=1840553; | | | |
| RA | Helm R., Strehler E.E.; | | | |
| RT | "Cloning an Escherichia coli gene encoding a protein remarkably | | | |
| RT | similar to mammalian aldehyde dehydrogenases."; | | | |
| RL | Gene 99:15-23(1991). | | | |
| RN | [4] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | SPECIES=S.flexneri; STRAIN=301 / Serotype 2a; | | | |
| RX | MEDLINE=22272406; PubMed=12384590; | | | |
| RA | Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H., | | | |
| RA | Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J., | | | |
| RA | Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S., | | | |
| RA | Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y., | | | |
| RA | Yu J.; | | | |
| RT | "Genome sequence of Shigella flexneri 2a: Insights into pathogenicity | | | |
| RT | through comparison with genomes of Escherichia coli K12 and O157."; | | | |
| RL | Nucleic Acids Res. 30:4433-4441(2002). | | | |
| RN | [5] | | | |
| RP | IDENTIFICATION. | | | |
| RC | SPECIES=E.coli; | | | |
| RA | Rudd K.E.; | | | |
| RL | Unpublished observations (AUG-1994). | | | |
| CC | -1- SIMILARITY: SOME, TO H.INFLUENZAE HI0659. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on use | | | |
| CC | by non-profit institutions as long as its content is in no way | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | |
| CC | or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL: AE000228; AAC74381.1; - | | | |
| DR | EMBL: D90768; BAAL1868.1; - | | | |
| DR | EMBL: D90767; BAAL1859.1; - | | | |

```

DR EMBL; M38433; -; NOT_ANNOTATED_CDS.
DR EMBL; AE015157; AAN42915.1; -.
DR PIR; F64878; F64878.
DR Ecogen; EG12431; YcJc.
DR InterPro; IPR007113; CypIn_sup.
DR InterPro; IPR001387; HTH_3.
DR Pfam; PF01381; HTH_3; 1.
DR SMART; SM00530; HTH_XRE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 185 AA; 20091 MW; D7D0D3FD794A7768 CRC64;

Query Match 4.7% Score 63.5; DB 1; Length 185;
Best Local Similarity 27.5% Pred. No. 2.9e+02;
Matches 33; Conservative 20; Mismatches 44; Indels 23; Gaps 6;

OY 145 NEAIPNDERPTTPVAMATTIRKLITGELLATASRQQLIDMHEADVACGPLLSALPAGW 204
DB 34 HSAITSTEDKVSFA--ISTLQILKLYGLSLSE-----FSEPEKPDFEYV----- 78
OY 205 FIADKSGAGCERGSRCITIALGPDCKPSRIIVV-----YTTGSOATMDER-NRQIAEIGASL 259
DB 79 -VINQDDLLIMGSGVSMKLVHNGNPNTLMIFFETQPGT--TTGERIKHQGEIGITVL 135

RESULT 13
HSLV_THEME
ID HSLV_THEME STANDARD; PRT: 176 AA.
AC Q9W1Z1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP-dependent protease hslv (EC 3.4.25.-).
GN HSLV OR TM0521.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uutterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith R.O., Venter J.C., Fraser C.M.;
RA *Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima";
RL Nature 399:323-329(1999).
CC -1- FUNCTION: Belongs to a proteasome-like degradation
CC complex (By similarity).
CC -1- SUBUNIT: INTERACTS WITH HSLU (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1B. HSLV SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001728; AAD35606.1; -.
DR PIR; G72365; G72365.
DR HSSP; P31059; 1E94.
DR MEROPS; T01_006; -.
DR TIGR; TM0521; -.
DR HAMAP; MF_00248; -; 1.
DR InterPro; IPR001353; Protease-protease.
DR Pfam; PF00227; proteasome; 1.
DR HydroLase; Protease; Complete proteome.
KW ACT_SITE 6 6
BT SIMILARITY.

```


| | | |
|----|---------------------------------------|--|
| DR | EMBL; Z30697; CAA83179.L; | -. |
| DR | PIR; J01931; J01931. | |
| DR | InterPro; IPR003875; Paramyxovir_NSC. | |
| DR | Pfam; PF02725; Paramyxo_NS_C; 1. | |
| KW | Nonstructural Protein. | |
| SO | SEQUENCE | 177 AA; 19926 MW; 76D8D46E4ED3FEB07 CRC64; |

| | | | | |
|-----------------------|------------------|--------------------|------------|-------------|
| Query Match | 4.88; | Score 65; | DB 1; | Length 177; |
| Best Local Similarity | 24.68; | Pred. No. 2.le+02; | | |
| Matches 51; | Conservative 17; | Mismatches 44; | Indels 34; | Gaps 5; |

| | | | |
|----|-----|--|------|
| OY | 93 | TVEELCSAATITMSHNTFAANLLLTGTGCKRELAFALHNMDQHYTRIDWEPELNEAIRPDE | 152 |
| | | :: | :: : |
| Db | 49 | TRISASHNSQQDOAKAACLAIVTKIDEENTAVNKS-----WEHSL----- | 90 |
| OY | 153 | KDTTTP-----VAMATTLRKLGLGELLITASROOILDMWEADKVAGPLIRSAIPA | 202 |
| | | : : | : |
| Db | 91 | ---VFPOCIAPRYSIIMFMITAVERKRLRESKMULTLSMFNQAL--MNVSK-SGEERNLRTA | 144 |
| OY | 203 | GWFIAID | 208 |
| Db | 145 | MMILAN | 150 |

```

OX NCBI_TaxID=2214;
RN
RP SEQUENCE FROM N.A. 35395 / DSM 2834;
RC STRAIN=C2A / ATCC 35395;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nisbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Althorpe D., Brown A.,
RA Linton L., McEwan P., McKernan K., Talama J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuelner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith R.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Metcalf W.M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RT "The genome of Methanocaldococcus jectus reveals extensive metabolic
RT and physiological diversity.";
RU Genome Res. 12:532-542(2002).
CC -1- CATALYTIC ACTIVITY: ATP + thymidine 5'-phosphate = ADP + thymidine
CC 5'-diphosphate.
CC -1- SIMILARITY: BELONGS TO THE THYMIDYLATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sid.ch).
CC -----
DR EMBL: AE011164; AM07774.1; -
DR HAMAP: MF_00165; -1.
DR InterPro: IPR000062; Thymidylate_kin.
DR Pfam: PF02223; Thymidylate_kin.1.
DR TIGRfam: TIGR00041; TDM_Kinase; 1.
DR PROSITE: PS01331; THYMIDYLATE_KINASE; FALSE_NEG.
DR Transferrase; Kinase; Nucleotide biosynthesis; ATP-binding;
KW Complete proteome.
FT NP_BIND 10 17 ATP (POTENTIAL).
SQ SEQUENCE 206 AA; 23303 MW; A19C2E58C85423B3 CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 206;
Best Local Similarity 18.7%; Pred. No. 2.3e+02;
Matches 41; Conservative 37; Mismatches 80; Indels 61; Gaps 10;

QY 56 LSRID-AGOEOLGRRIHYSQNDLVESPTVEKHLTDSMTVRELCSAATMSDNTANLIL 114
DB 8 LEGIDSGCKSTVAKLQ-ENSELRVFEVPTREPTRC-----TLGNVAVENAIQ 55
QY 115 TTIGPKELTALHNMGDHYTRLDRWEPELNEAIPNDRDTTPVAMATTLKLTGELL 174
DB 56 SDTQDLAELEFTADHAELAKLVKPALEGGKTVISDRYSDSRYAOGITTLKNRLDNL 114
QY 175 TLASRQQLIDMEADKVAAPLRSALPAGW-----FIAD-----KSC-AGE----- 214
DB 115 -----EMVR-----DLHKGTVIPDLTFLFDEIEPIAVKRGCKRGEQTKF 154

QY 215 -----RGSRTIALGPGKPSRIIVITYTGSQATDE 247
DB 155 EKIEFLRGVRELFLGLAAE-EPERFVIVDASGSPEDVEK 192

RESULT 9
VNSC_RINDK STANDARD; PRT; 177 AA.
AC P35948;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Nonstructural protein C.
OS Rinderpest virus (strain Kabete O) (RDV).

```

```

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID=11242;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92410646; PubMed=1529555;
RA Yamanaka M., Dale B., Crisp T., Cordell B., Grubman M., Ylisma T.;
RT "Sequence analysis and editing of the phosphoprotein (P) gene of
RT rinderpest virus.";
RU Virology 190:553-556(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S44819; AB23269.1; -
DR PIR: B43387; B43387.
DR InterPro: IPR003875; Paramyxovir_NSC.
DR Pfam: PF02725; Paramyxo_NS_C; 1.
KW Nonstructural protein.
SQ SEQUENCE 177 AA; 19926 MW; 116971B140A39F11 CRC64;

Query Match 4.8%; Score 65; DB 1; Length 177;
Best Local Similarity 24.6%; Pred. No. 2.1e+02;
Matches 31; Conservative 17; Mismatches 44; Indels 34; Gaps 5;

QY 93 TVRELCSAATMSDNTANLILTTIGPKELTALHNMGDHYTRLDRWEPELNEAIPNDE 152
DB 49 TIRISASHASOODLQAKACLAIVTRIDLEATVMS9-----WEHSL----- 90
QY 153 RDTTTP-----VAMATTLKLTGELLTLASRQQLIDMEADKVAAPLRSALPA 202
DB 91 --VTPGCIAPRYSTIMEITAVKRLRESKMLTSLSPNAL--MMVSK-SGEEMRLRPA 144

QY 203 GMFIAD 208
DB 145 MMLIAN 150

RESULT 10
VNSC_RINDK STANDARD; PRT; 177 AA.
AC 003339;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Nonstructural protein C.
GN C.
OS Rinderpest virus (strain RBOK) (RDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbilliviruses.
OX NCBI_TaxID=36409;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=93155658; PubMed=8429304;
RA Barron M.D., Shalla M.S., Barrett T.;
RT "Cloning and sequence analysis of the phosphoprotein gene of
RT rinderpest virus.";
RU J. Gen. Virol. 74:299-304(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68311; CAA48391.1; -

```

```
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome-binding factor A.
GN RBFA OR U0321.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3.
RA MEDLINE=20500219; PubMed=11048724.
RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Casella G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RT
RL Nature 407:757-762(2000).
CC -1- FUNCTION: Associates with free 30S ribosomal subunits (but not
CC with 30S subunits that are part of 70S ribosomes or polysomes).
CC Essential for efficient processing of 16S rRNA. May interact with
CC the 5' terminal helix region of 16S rRNA (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE RBFA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AE002129; AAF0730.1; -
DR HAMAP: MF_000003; -- 1
DR InterPro: IPR000238; Rib_bind_facta.
DR Pfam: PF02033; RBFA. 1.
DR ProDom: PD007327; Rib_bind_facta; 1.
DR TIGRFAMs: TIGR00082; rbfA. 1.
DR PROSITE: PS01319; RBFA; FALSE_NEG.
KW rRNA processing; Complete proteome.
SQ SEQUENCE 116 AA; 13247 MW; A/9DCC71F0547514 CRC64;

Query Match 4.9%; Score 66.5; DB 1; Length 116;
Best Local Similarity 28.9%; Pred. No. 91;
Matches 24; Conservative 16; Mismatches 24; Indels 19; Gaps 5;

QY 144 LINEAIPNDERDTTPVAMATTLRLKLTGEL-----LTLASRQQLIDWME-ADRVAGPL 196
DB 18 INNALEINDIKTAKLRVAVR--LSNDLSYAKITFLDHRKESMLKYLENKKVSG-L 74
QY 197 RSLAPAGW-----FIADKS 210
DB 75 RSKLAEMWTSYKVPBELRFVIDET 97

RESULT 7
OGGI_ARCFU STANDARD; PRT; 198 AA.
AC 029876;
DT 16-OCT-2001 (Rel. 40, Last Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable N-glycosylase/DNA lyase [includes: 8-oxoguanine DNA
DE glycosylase (EC 3.2.2.-); DNA-(apurinic or apyrimidinic site) lyase
DE (EC 4.2.99.18) (AP lyase)].
GN OGG OR AF0371.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
```

```
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RT
RL Nature 390:364-370(1997).
CC -1- FUNCTION: Responsible for removing an oxidatively damaged form of
CC guanine (7,8-dihydro-8-oxoguanine - 7-oxoG) from DNA. Also nicks
CC DNA at apurinic/aprimidinic sites (AP sites) (By similarity).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SIMILARITY: BELONGS TO THE OGG1 FAMILY 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AE001079; AAB90876.1; -
DR PIR: C69296; C69296.
DR TIGR: AF0371; -
DR HAMAP: MF_00241; -- 1.
DR InterPro: IPR003265; Endo_3c.
DR Pfam: PF00730; Hnh-GPD. 1.
DR SMART: SM00478; ENDO3c; 1.
DR SMART: SM00478; ENDO3c; 1.
KW Hydrolyase; Nuclease; Endonuclease; Lyase; DNA repair; Glycosidase;
KW Multifunctional enzyme; Complete proteome.
FT ACT-SITE 122 122 BY SIMILARITY.
SQ SEQUENCE 198 AA; 22639 MW; 3A5C033AA12F3FFB CRC64;

Query Match 4.9%; Score 65.5; DB 1; Length 198;
Best Local Similarity 25.7%; Pred. No. 2.2e+02;
Matches 37; Conservative 19; Mismatches 53; Indels 35; Gaps 8;

QY 14 QGARGYIELDINSGRILSEFRPEEPMASTFKVLGAVL-----SRIDAG---OEOIG 67
DB 15 QGKER-GEVEFD-----FRPFDSEVATITBELAFCLSTANSSATAGLKFORLLG 64
QY 68 RRIHYSQ-----NDLVEYSPYTERKHLDTGVTRELCSPAATMSDNTANLLITTT 117
DB 65 QGVGVKEALTLIAGVRFHNRKAEY-----IREAFKSRIVKALEAESKREILITK 117
QY 118 G-GPKELTAPLHMNG-DHYTRIDR 139
DB 118 GLGMRKASHPLRNVRGREDVAIIDR 141

RESULT 8
KTHY_METAC STANDARD; PRT; 206 AA.
AC 087859;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable thymidylate kinase (EC 2.7.4.9) (dTMP kinase).
GN TMK OR MA4433.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
```

QY 100 AATMSDNTAANLLTTIGPKR-----LTAF-LHNMGDHYRLDMEPE 143
 DB 45 ANVAV-DGSTVNLGLMFDAGQEDYNRLPLSYRGADVFLALSLISRASENISKWPE 103
 QY 144 LNEAIPNDERDTPVAMATTLKLLTGELLTL-ASRQOLIDMEADKVA--GPLLBSA 199
 DB 104 LRAHAP-----TVPI-----VLVGTRLDREDRQYILDHGCAFPITTAAGSELKKA 149
 QY 200 LPACWFI 206
 DB 150 IGAAYVL 156
 RESULT 4
 ID NDK_TREPA STANDARD; PRT; 149 AA.
 AC 083974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
 GN NDK OR TP1010.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RA ERASER C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T.,
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Harch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT *Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete.";
 RL Science 281:375-388(1998).
 CC -1- FUNCTION: MAJOR ROLE IN THE SYNTHESIS OF NUCLEOSIDE TRIPHOSPHATES
 CC OTHER THAN ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + nucleoside diphosphate = ADP +
 CC nucleoside triphosphate.
 CC -1- SUBUNIT: Homotrimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the NDK family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001269; AAC65961.1; -.
 DR PIR: F71252; F71252.
 DR HSSP: P22887; INPK.
 DR TIGR: TP1010; -.
 DR HAMAP: MF_00451; -; 1.
 DR InterPro: IPR001564; NDK.
 DR Pfam: PF00334; NDK.1.
 DR PRINTS: PRO1243; NUCDPKINASE.
 DR PRODOM: PD001018; NDK.1.
 DR SMART: SM00562; NDK.1.
 DR PROSITE: PS00469; NDP_KINASES; 1.
 KW Transferase; Kinase; ATP-binding; Complete proteome.
 FT ACT_SITE 117 117 By SIMILARITY.
 SQ SEQUENCE 149 AA; 16657 MW; BB329539671E80BB CRC64;
 Query Match 5.0%; Score 68; DB 1; Length 149;
 Best Local Similarity 21.6%; Pred. No. 95;

Matches 35; Conservative 25; Mismatches 64; Indels 38; Gaps 7;
 QY 29 GILSEFRPEERFPMMSTFKVLLCGAVLSRIDAGQEDLGRIRHSQNDLVESPYTEKHL 88
 DB 21 GEVLSRF--ERKGLVTLRLC-----VDATATL---HYAEHKEKFPYPSLIAYI 67
 QY 89 TDGMTVRELCSAATMSDNTAANLLTTIGPRELTFALHNMGDHYRLDMEPEPLNEAI 148
 DB 68 TSAPV-----ALAFKGNALSLVRLCGS-----TRVERAQP---GTI 103
 QY 149 PND-ERDPTTPVAMATTLKLLTGELLTLASRQOLIDMEAD 189
 DB 104 RGDFAIRTTTNTIVHSDPESARALALYFSADDFVEMRGN 145
 RESULT 5
 ID HC3L_THIFE STANDARD; PRT; 122 AA.
 AC P80509;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Cytochrome-c3 hydrogenase, large chain (EC 1.12.2.1) (Hydrogenase)
 DE (Fragments).
 GN HOKG.
 OS Thioacillus ferrooxidans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Acidithiobacillales;
 OC Acidithiobacillaceae; Acidithiobacillus.
 OX NCBI_TaxID=920;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=ATCC 19859;
 RA MEDLINE=96241862; PubMed=8661919;
 RA Fischer J., Quentemeier A., Kostka S., Kraft R., Friedrich C.G.;
 RT *Purification and characterization of the hydrogenase from
 RT Thioacillus ferrooxidans.";
 RL Arch. Microbiol. 165:289-296(1996).
 CC -1- CATALYTIC ACTIVITY: H(2) + 2 ferricytochrome c3 + 2
 CC ferrocyclochrome c3.
 CC -1- COFACTOR: IRON.
 CC -1- CAUTION: THE ORDER OF THE SUBMITTED PEPTIDES IS UNKNOWN. IT IS
 CC ALSO UNKNOWN IF GAPS EXIST BETWEEN THE PEPTIDES.
 CC -----
 CC Oxidoreductase: Iron.
 KW FT NON_CONS 20 21
 FT NON_CONS 29 30
 FT NON_CONS 35 36
 FT NON_CONS 42 43
 FT NON_CONS 59 60
 FT NON_CONS 72 73
 FT NON_CONS 78 79
 FT NON_CONS 87 88
 FT NON_CONS 98 99
 FT NON_CONS 107 108
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 13146 MW; EB4FAF365348DA0B CRC64;
 Query Match 5.0%; Score 67; DB 1; Length 122;
 Best Local Similarity 26.4%; Pred. No. 89;
 Matches 28; Conservative 13; Mismatches 23; Indels 42; Gaps 4;
 QY 19 VGYIELDNSGKILSEFRPERFPMMSTFKVLLCGAVLSRIDAGQEDLGRIRHSQNDLV 78
 DB 12 VGRVEDDLSXSLLEFR-----NAILARFGG-----LG 41
 QY 79 EYSPYTEKHLTDGMTVRELCSAATMSDNTAANLLTTIGPKREL 124
 DB 42 KYAPFTGTNYELGV-----ISGDKDQAGLVLT---PREST 75
 RESULT 6
 ID RBFA_UREPA STANDARD; PRT; 116 AA.
 AC Q9FQH0;

| | | | | |
|-----------------------|--------|---------------|-------|-------------|
| Query Match | 5.68; | Score 75; | DB 1; | Length 201, |
| Best Local Similarity | 22.68; | Pred. No. 38; | | |

```
QY      44 MSTFEVLGAVLSRIDAGEQLGR---IHSNDL-VESPYTEKHITDGTVELCS 99
          ||| : | : : : ||| : | : | : ||| : |
Db      1 MSTARFIC-----VTYGDAVGAKNCMLSYSTNTPETDVPE-----TYFDNS 44
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51; Search time 5.57143 Seconds
(without alignments)
2219.902 Million cell updates/sec

Title: SEQ2_157T
Sequence: 1348
1 HPETLVKVKAEQDQGVAVG.....TMDERNRQIAEIGASLIKHW 263

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 42617

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|-------------|---------------------|
| 1 | 77 | 5.7 | 191 | YDA4_SCHPO | Q10346 schizosacch |
| 2 | 75 | 5.6 | 201 | RAC2_DICDI | Q99P80 dictyostell |
| 3 | 68.5 | 5.1 | 196 | RAC2_LOTJA | Q40220 lotus japon |
| 4 | 68 | 5.0 | 149 | NDK_TREPA | O83974 treponema p |
| 5 | 67 | 5.0 | 122 | HC3L_THIFE | P80509 thibacilli |
| 6 | 66.5 | 4.9 | 116 | RBRF_UREPA | Q9PQ80 ureaplasma |
| 7 | 65.5 | 4.9 | 198 | OGGI_ARCFU | Q28876 archaeoglob |
| 8 | 65.5 | 4.9 | 206 | KTHI_METAC | O83H93 methanosarc |
| 9 | 65 | 4.8 | 177 | VNSC_RINDK | P33948 rinderpest |
| 10 | 65 | 4.8 | 177 | VNSC_RINDR | O03339 rinderpest |
| 11 | 63.5 | 4.7 | 144 | MARR_ECOLI | P27245 escherichia |
| 12 | 63.5 | 4.7 | 185 | YCUC_ECOLI | P38522 escherichia |
| 13 | 63 | 4.7 | 176 | HSIV_THEMA | Q9WY21 thermotoga |
| 14 | 62.5 | 4.6 | 150 | PDIV_SALTY | O9JDM6 salinimella |
| 15 | 62.5 | 4.6 | 172 | YDEI_ECOLI | P31131 escherichia |
| 16 | 62.5 | 4.6 | 182 | PYRE_STRGO | O9X817 streptomyce |
| 17 | 62.5 | 4.6 | 195 | TRPF_THREVO | Q979V6 thermoplasma |
| 18 | 62.5 | 4.6 | 200 | RR4_PELINE | O9FSD9 pelilla nees |
| 19 | 62 | 4.6 | 174 | IHHB_RAP | P17491 rattus norv |
| 20 | 62 | 4.6 | 182 | Y861_BIFLO | O85520 bifidobacte |
| 21 | 62 | 4.6 | 186 | YCER_SALTY | P40822 salinimella |
| 22 | 61.5 | 4.6 | 178 | HSIV_RALSO | O8J3D7 raistonia s |
| 23 | 61 | 4.5 | 121 | SECR_HUMAN | P09683 homo sapien |
| 24 | 61 | 4.5 | 200 | TATB_CAUCR | O966C1 caulobacter |
| 25 | 61 | 4.5 | 202 | COAT_ELV | P35927 erysimum la |
| 26 | 60.5 | 4.5 | 103 | RS10_NERGO | P48851 neisseria g |
| 27 | 60.5 | 4.5 | 146 | HBG_NABIT | P02099 oryctolagus |
| 28 | 60.5 | 4.5 | 156 | BFR_AZOV1 | P22759 azotobacter |
| 29 | 60.5 | 4.5 | 160 | TATB_HELPY | O25700 heliobacte |
| 30 | 60.5 | 4.5 | 184 | MLRI_SCHPO | O9UN95 schizosacch |
| 31 | 60.5 | 4.5 | 193 | SAR2_LYCES | P52884 lycopersico |
| 32 | 60.5 | 4.5 | 196 | ALKH_BACSU | P50846 b kmg/kdpg |
| 33 | 60 | 4.5 | 148 | CALM_BLAEM | O9niy6 blastoclad |

| | | | | | | |
|----|------|-----|-----|---|------------|--------------------|
| 34 | 60 | 4.5 | 159 | 1 | GRE4_CHLTE | O8KCH5 chlorobium |
| 35 | 60 | 4.5 | 161 | 1 | PIN_BPT4 | P07068 bacterioph |
| 36 | 60 | 4.5 | 178 | 1 | PYRE_ARCFU | O28533 archaeoglob |
| 37 | 60 | 4.5 | 178 | 1 | UCRI_ANASP | P70758 anabaena sp |
| 38 | 60 | 4.5 | 184 | 1 | ARL2_DROME | Q06849 drosophila |
| 39 | 60 | 4.5 | 205 | 1 | RS2_AERPE | O9YB45 aeropyrum p |
| 40 | 59.5 | 4.4 | 103 | 1 | RS10_NEIMA | O9JF21 neisseria m |
| 41 | 59.5 | 4.4 | 133 | 1 | CCRN_PAROL | O57312 paratrichy |
| 42 | 59.5 | 4.4 | 184 | 1 | HRPL_PSEST | P37929 pseudomonas |
| 43 | 59.5 | 4.4 | 196 | 1 | I196_ASFB7 | P27943 african swi |
| 44 | 59.5 | 4.4 | 198 | 1 | UPP_ARATH | O65583 arabidopsis |
| 45 | 59.5 | 4.4 | 205 | 1 | RS4_ECOLI | P02354 escherichia |

ALIGNMENTS

RESULT 1
YDA4_SCHPO
ID YDA4_SCHPO STANDARD: PRT; 191 AA.
AC Q10346;
DT 01-OCR-1996 (Rel. 34, Created)
DT 01-OCR-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ClpF12.04c in chromosome I.
GN SPAC1F12.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_Taxid:4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RC MEDLINE-21848401, PubMed-11859360;
RA Wood V., Gwilliam R., Rajendram M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayes J., Baker S., Baaham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Mblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Stimmings M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkert G., Aert R., Robben J., Grymoprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritsc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler S., Gloux S., Lelaure V., Motlier S.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McComble W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nuzre P.;
RT *The genome sequence of Schizosaccharomyces pombe.*;
RL Nature 415:871-880(2002).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL: 269944; CAA93808.1; -
CC PIR: T38062; S67447. -
CC GenedB_Spombe: SPAC1F12.04C; -

THIS PAGE BLANK (USPTO)

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0825

C:Superfamily: Bacillus subtilis hypothetical protein ykka

Query Match 5.3%; Score 71; DB 2; Length 177;

Best Local Similarity 20.3%; Pred. No. 99;

Matches 38; Conservative 28; Mismatches 57; Indels 64; Gaps 8;

42 PMATFKVLLCGAVSRIDAGQEQGRRIHYSQNDLVEYSPYTERKHLTDGTVRELCSAA 101

11 PEMDTSVGLFYAMVRENDV-----RLH-----HLIEDVTEELYYKG 47

102 ITMSDNTAANLLTITIGPKELTAPLHMGDHTRLD-RKEPEL-NKAIPND----- 151

48 SDGDENSKAQL-----NHLTYDVAVRVEFKGKALPDLSEAEHSGPM 89

152 -ERDTTPVAMATTLRLKLTGELTLASHQ-----LIDMEADKVAQPLRSALP 201

90 VDKDKGLPVVTSLSVQELIEKORYVALIKERCQALHDDDLAKRIPYEEERQATIRMG-L 148

202 AGNFIAD 208

149 --WRMAD 153

RESULT 11

683096 conserved hypothetical protein PA4399 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 24-May-2001

C:Accession: G83096

R:Stover, C.K.; Pham, X.Q.; Ertvin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B. adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim, .; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Accession: G83096

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-192 <STO>

A:Cross-references: GB:AE004855; GB:AE004091; NID:g9950621; PIDN:AA607787.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4399

C:Superfamily: conserved hypothetical protein DR1638

Query Match 5.3%; Score 71; DB 2; Length 192;

Best Local Similarity 26.0%; Pred. No. 1.1e+02;

Matches 53; Conservative 22; Mismatches 87; Indels 42; Gaps 9;

53 GAVLSRI-----DAGQEQ--GRIHYSQNDLVEYSPYTERKHLTDGTVRELCSA----- 100

2 GNLSTIYRTGGRGFIAGGRVPSKPRIEAIGAVDELSSQGLLAELELARGAMP 61

101 -----ATMSDNTAANLLTITIGPRELTAFLHMGDHTRL-----DRKEPEL-----N 145

62 GLEIYQALAPVQHR-----LEDIGELAMPYRALDEFEVARLESCIDRMDELGRN 116

146 EAIPINERDTTVAAMATTLKLTGELTLASRQQLIDMEADKVAQPLRSALPAGWF 205

117 FILPGGSR-----PVAQAHVCRSLAR-----SAERRCOALDQETLEGGVIRYLRLSDLLF 168

206 IADKSGAGGSRGII--AALGPD 227

169 VARAIARQGVAEILMEAAKPD 192

RESULT 12

884217

hypothetical protein Vng0594h [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: E84217

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky

; Leitner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja

proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: E84217

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-108 <STO>

A:Cross-references: GB:AE004437; NID:g10580187; PIDN:AA619105.1; GSPDB:GN00138

C:Genetics:

A:Gene: Vng0594H

Query Match 5.2%; Score 70; DB 2; Length 108;

Best Local Similarity 27.5%; Pred. No. 61;

Matches 28; Conservative 16; Mismatches 42; Indels 16; Gaps 3;

44 MSTFKVLLCGAVSRIDAGQEQGRRIHYSQNDLVEYSPYTERKHLTDGTVRELCSAAT 103

1 MDVTYELDGLALRLAETD-----RVFEVRFDALAEVTVTLFRHGDGDRV-----GSIT 50

104 MSDNTAANLLTITIGPREL-----TAPLHMGDHTRLDR 139

51 NDDGTDRTMARLTVPGSDFTAVEPTSFVAIVDAATRTDR 92

RESULT 13

AB0192

conserved hypothetical protein YP01575 [imported] - Yersinia pestis (strain C092)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002

C:Accession: AB0192

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M

deno-Parraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AB0192

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-170 <KUR>

A:Cross-references: GB:AL50842; PIDN:CAC90397.1; PID:g15979615; GSPDB:GN00175

C:Genetics:

A:Gene: YP01575

C:Superfamily: Escherichia coli hypothetical protein b0354

Query Match 5.2%; Score 70; DB 2; Length 170;

Best Local Similarity 21.0%; Pred. No. 1.1e+02;

Matches 22; Conservative 25; Mismatches 44; Indels 14; Gaps 3;

56 LSRIDAGQEQGRRIHYSQNDLVEYSPYTERKHLTDGTVRELCSAATMSDNTAANLLT 115

70 MKKIDISKNGIG--FNFTDNNLI-----KKIYVDKLTQAOALISRLAIA-----RLVV 115

116 TIGPRELTAFLHMGDHTRLDRKEPELNAIPNDESDTTTPVA 160

116 DNGSESEFVAITPASVADKIQORDASSIVLNSALSOEDGDEPYA 160

RESULT 14

F84251

hypothetical protein Vng0953c [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: F84251

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky

; Leitner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja

Jung, K.H.; Alam, M.; Freitas, T.

Best Local Similarity 22.6%; Pred. No. 34;
Matches 43; Conservative 22; Mismatches 69; Indels 56; Gaps 7;

OY 104 MSNTANALLTTGGKELTAFLNMGDHVRRLDMEPELNEAIPNDERDTTPVAMAT 163
Db 1 MSTWHSIMLMQNSGIDKIAAILVN---VARLD-----PASSKSTQVSMIN 46
OY 164 TLRLKLTGELLTLASROOLDIMMEADKVAAGPLRSALPAGWFIADKSGAGERSGIIAA 223
Db 47 EPRC-----ILRLPGLKLIYNFRKDSPEEYMSNAINGYYTE--GLAFLGGKQIISI 99
OY 224 LCPDGRP-----SRIVYITGSGQATMDERNROI----- 252
Db 100 ---SKPLEDKLMMSRFMLDLTLITVQLRKRETEDEKHOOLDIASNLASLPLCTHWS 155
OY 253 AETGASLIKH 262
Db 156 VENGAGLIKH 165

RESULT 3

hypothetical protein u1756t [Imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45195
R:Robison, K.
submitted to the EMBL Data Library, September 1994
A:Reference number: Z16911
A:Accession: T45195
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-113 <KEI>
A:Cross-references: EMBL:U15180; PIDN:AAA62885.1

Query Match 5.5%; Score 74.5; DB 2; Length 113;
Best Local Similarity 28.6%; Pred. No. 27;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;

OY 130 MDDHV--TLRDMKEPELN---EAIPNDERDT---TFPVAMATTLRLKLTGELLTLASRQ 180
Db 1 MCDIGMEREGREWTGTCPLRVPGDEPTLDGRASPEDLIIT--NLSPTIMSGPPPS 58
OY 181 QLIDMMEA-DKVAGPLL-----KSALPAGWFIADKSGAGERSGIIAALGPDPKPSR 232
Db 59 RDDWVEPPDALGTVAFATGDKATMPAVAGI---GASTRGS-GILASLSPFRQPAR 112

RESULT 4

AD3606

molybdopterin biosynthesis mog protein [Imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AD3606
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, M.; Mazur, M.; Coleman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD33252; PMID:11756688
A:Accession: AD3606
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-172 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL54015.1; PID:917984966; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMET10773
A:Map position: 11
C:Superfamily: molybdenum cofactor biosynthesis protein B moab

Query Match 5.5%; Score 74.5; DB 2; Length 172;
Best Local Similarity 25.5%; Pred. No. 48;
Matches 40; Conservative 14; Mismatches 42; Indels 61; Gaps 9;

OY 83 VTEKHLTDGM-TVELCSAAITMSDNTAANLLTTIG-CPKELTAFLNMGDHVRRLDRLW 140
Db 39 IYRVRIIPDGMEVRD---TLIDCDTVACDLITLTTGGTSPS----- 76

OY 141 EPLNEAIPNDERDTTPVAMATTLRLKLT--GELLTLASROOLDIMMEADKVAAGPLRS 198
Db 77 -----PRDE-----TPKAMKAVLHKELPGEGQMRVRSLEQ----- 107
OY 199 ALPAGWFIADKSGAGERSGIIAALPDPKPSRIYV 235
Db 108 -TPFA--VLSRQTASRGKSPF---LMLPGKPASIAM 138

RESULT 5

T36984

hypothetical protein SCJ11.13 - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
C:Accession: T36984
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T36984
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-152 <OLI>
A:Cross-references: EMBL:AL109949; PIDN:CAB52898.1; GSPDB:GN00070; SCODEB:SCJ11.13
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SCJ11.13
C:Superfamily: Streptomyces coelicolor hypothetical protein SCJ9A.17

Query Match 5.3%; Score 71.5; DB 2; Length 152;
Best Local Similarity 23.4%; Pred. No. 73;
Matches 37; Conservative 16; Mismatches 50; Indels 55; Gaps 8;

OY 106 DNTAANLLTTTGGKELTAFLNMGDHVRRLDMEPELNEAIPNDERDTTPVAMATTL 165
Db 25 DQAAARTLAQVPTPPADLAFYESTIGD-VT---WEDVNGYFLNP----- 65
OY 166 RLTLGELLTLASROOLDIMMEADKVAAGPLRSALPAGWFIADKSGAGERSGIIAALG 225
Db 66 ---AGDILLRLQEVGVVDGTDEKSRGLVI-----GSNGG-GLIYVAG 104
OY 226 PDGKPSRIYVITGSGQATMDERN-----ROIARI 255
Db 105 PDG-----AVYRTRT-ASIDFAELDKRVADGLRQFLLEL 135

RESULT 6

T09136

ADP-ribosylation factor homolog ARL3 - Trypanosoma brucei

N:Alternate names: ADP ribosylation factor 3 homolog
C:Species: Trypanosoma brucei
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Feb-2001
C:Accession: T09136
R:Brinsud, F.; Verreume, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, M.; Blochem, Parisitol. 94, 249-264, 1998
A:Title: Conserved organization of genes in trypanosomatids.
A:Reference number: Z16580; MUID:98418771; PMID:9747975
A:Accession: T09136
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-190 <BRI>
A:Cross-references: EMBL:AF031926; NID:g3452211; PIDN:AAC32774.1; PID:g3452215
A:Experimental source: strain Antrati
C:Note: small G-protein
C:Genetics:
A:Gene: ARL3
C:Superfamily: ADP-ribosylation factor
C:Keywords: blocked amino end; lipidprotein; myristylation; nucleotide binding; P-loop
F:24-31/Region: nucleotide-binding motif A (P-loop)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:19:51 : Search time 10.2857 seconds
(without alignments)
2458.980 Million cell updates/sec

Title: SEQ2_157T

Perfect score: 1348

Sequence: 1 HPETLVKVKDAEDQLGARVQ.....TMDERRRQIAETGASLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 105605

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 524 | 38.9 | 105 | 2 | JC2566 |
| 2 | 77 | 5.7 | 191 | 2 | S67447 |
| 3 | 74.5 | 5.5 | 113 | 2 | T45195 |
| 4 | 74.5 | 5.3 | 152 | 2 | AD3606 |
| 5 | 71.5 | 5.3 | 190 | 2 | T36984 |
| 6 | 71.5 | 5.3 | 195 | 2 | T09136 |
| 7 | 71.5 | 5.3 | 195 | 2 | T36975 |
| 8 | 71.5 | 5.3 | 200 | 2 | G97064 |
| 9 | 71 | 5.3 | 153 | 2 | E81708 |
| 10 | 71 | 5.3 | 177 | 2 | A83753 |
| 11 | 71 | 5.3 | 192 | 2 | G83096 |
| 12 | 70 | 5.2 | 108 | 2 | E84217 |
| 13 | 70 | 5.2 | 170 | 2 | AB0192 |
| 14 | 68.5 | 5.1 | 145 | 2 | P84251 |
| 15 | 68.5 | 5.1 | 167 | 2 | D87360 |
| 16 | 68.5 | 5.1 | 192 | 2 | A83587 |
| 17 | 68 | 5.0 | 149 | 2 | F71252 |
| 18 | 67.5 | 5.0 | 131 | 2 | AD2281 |
| 19 | 67.5 | 5.0 | 184 | 2 | T21126 |
| 20 | 66.5 | 4.9 | 116 | 2 | C82906 |
| 21 | 66.5 | 4.9 | 177 | 2 | D90227 |
| 22 | 66.5 | 4.9 | 198 | 2 | D95285 |
| 23 | 65.5 | 4.9 | 42 | 2 | F56978 |
| 24 | 65.5 | 4.9 | 181 | 2 | AB1902 |
| 25 | 65.5 | 4.9 | 195 | 2 | AE0623 |
| 26 | 65.5 | 4.9 | 198 | 2 | C69296 |
| 27 | 65 | 4.8 | 148 | 2 | E75283 |
| 28 | 65 | 4.8 | 150 | 2 | T08585 |
| 29 | 65 | 4.8 | 160 | 1 | E69186 |

| | | | | | | |
|----|------|-----|-----|---|--------|---------------------|
| 30 | 65 | 4.8 | 168 | 2 | B75498 | conserved hypothet |
| 31 | 65 | 4.8 | 177 | 1 | B43387 | nonstructural prot |
| 32 | 65 | 4.8 | 177 | 2 | J01931 | nonstructural prot |
| 33 | 65 | 4.8 | 180 | 2 | C71869 | hypothetical prote |
| 34 | 65 | 4.8 | 180 | 2 | G70912 | hypothetical prote |
| 35 | 64.5 | 4.8 | 145 | 2 | H75262 | hypothetical prote |
| 36 | 64.5 | 4.8 | 151 | 2 | D81333 | probable protein-t |
| 37 | 64 | 4.7 | 113 | 2 | D70580 | hypothetical prote |
| 38 | 64 | 4.7 | 162 | 2 | AG0769 | probable acetyltra |
| 39 | 64 | 4.7 | 178 | 2 | 1K0124 | outer surface prote |
| 40 | 64 | 4.7 | 191 | 2 | E95333 | hypothetical prote |
| 41 | 63.5 | 4.7 | 164 | 2 | H82336 | regulator of sigma |
| 42 | 63.5 | 4.7 | 166 | 2 | C90029 | hypothetical prote |
| 43 | 63.5 | 4.7 | 168 | 2 | T20606 | hypothetical prote |
| 44 | 63.5 | 4.7 | 179 | 2 | AB1994 | hypothetical prote |
| 45 | 63.5 | 4.7 | 185 | 2 | D90863 | hypothetical prote |

ALIGNMENTS

RESULT 1

JC2566
bla protein - Pseudomonas aeruginosa (fragment)

C:Species: Pseudomonas aeruginosa

C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 03-May-1996

C:Accession: JC2566

R:West, S.E.H.; Schweizer, H.P.; Dall, C.; Sample, A.K.; Runyen-Janecky, L.J.
Gene 148(128), 81-86, 1994

A:Title: Construction of Improved Escherichia-Pseudomonas shuttle vectors derived fro

A:Reference number: JC2565

A:Note: due to a typographical error the volume number 148 appears as 128

A:Accession: JC2566

A:Molecule type: DNA

A:Residues: 1-105 <WES>

C:Genetics:

A:Gene: bla

C:Superfamily: beta-lactamase I

Query Match 38.9% Score 524; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 2.9e-37;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 160 AMATTLRRLTGGELLTASRQOLIDMNEADKVAQPLRSALPAGWFTADKSGAGERSRG 219
|||||
Db 2 AMATTLRRLTGGELLTASRQOLIDMNEADKVAQPLRSALPAGWFTADKSGAGERSRG 61
OY 220 IIAALGPDGKPSRIVYITTSQATMDERNRQIAETGASLIKHW 263
|||||
Db 62 IIAALGPDGKPSRIVYITTSQATMDERNRQIAETGASLIKHW 105

RESULT 2

S67447
hypothetical protein SPAC1F12.04c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C:Accession: T38062; S67447

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, March 1996

A:Reference number: T38062

A:Accession: T38062

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-191 <MC2>

A:Cross-references: EMBL:569944; NID:91217974; PIDN:CAA93808.1; PID:91217976; GSPDB:G

A:Experimental source: strain 97zh-; cosmid c1f12

C:Genetics:

A:Gene: SPAC1F12.04c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC1F12.04c

Query Match 5.7% Score 77; DB 2; Length 191;

```
RESULT 15
US-10-156-761-12656
; Sequence 12656, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMODA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12656
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12656

Query Match      4.88; Score 64.5; DB 15; Length 195;
Best Local Similarity 22.1%; Pred. No. 1.4e+02;
Matches 50; Conserved 21; Mismatches 88; Indels 67; Gaps 9;

QY      16 GARVGYIELDNSGKI-----LESFRPEERFPMSTFKVLGCA-----VLSRIDAG-- 62
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      4 GTATASTELRDSRGICLDLATTPHEERLDSADRLRAMATGSGLVPPGTPLEPADASWP 63
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      63 -----QEQLGRRIRHYSQNDLVEYSPYTERKHLTDGWTVAELCSAATMSDNTAANLLTT 116
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      64 VRFRELNGQIGQLVHGE-----LEGKPY-----DAILARVNAVA 97
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      117 IGGPKELTAPLHNNGDHVTRLDRWEPELNEALPNDERDTPPYAMATTUKLLTGELITL 176
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      98 IGSPPPLCAVRADGVLVRLPA-APCGCELLAATARDTV-----ELLTD 141
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY      177 ASRQQLIDWMEADKVAPPLRSALPAG---WFIADKSGAGERGSR 218
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB      142 PAARASIRHCEGDMC--PLVYLDTSRGRRRRWCSETCGNRRERVAR 185
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
```

Search completed: September 10, 2003, 12:33:18
Job time : 17.5714 secs

```

Db      61SELFPAEAERGELNFGAGYILEVSTPGVDNPL---TLPRHW-----RRNRGRLLVA 107
QY      224 LGPDGNP--SRIVVITYTGSQAIMDERNRQIAEIGASLIKH 262
      1 111 :11 : :111::1 :1
Db      108 LDQDGKKRVARIGALNDAETHVHLERRKKILEVYTLLELHA 148

RESULT 12
US-10-156-761-8623
: Sequence 8623, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 8623
: LENGTH: 194
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-8623

Query Match      5.0%; Score 67; DB 15; Length 194;
Best Local Similarity 38.6%; Pred. No. 79;
Matches 17; Conservative 9; Mismatches 12; Indels 6; Gaps 2

QY      202 AGCWFIADKSGAGERSRG---ITAIAGPDG--KPSRIVIVYTT 239
      11: :111111:1 11111:111:1:1:1
Db      75 SGRVWPRSGAGERTGCAEYRITAIIGPLAVHEPVRVAVVTT 118

RESULT 13
US-10-127-816-11
: Sequence 11, Application US/10127816
: Publication No. US20030104416A1
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Fox, Brian A.
: APPLICANT: Klucher, Kevin M.
: APPLICANT: Taft, David W.
: APPLICANT: Kindsvogel, Wayne R.
: TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY
: FILE REFERENCE: 01-17
: CURRENT APPLICATION NUMBER: US/10/127,816
: CURRENT FILING DATE: 2002-04-19
: PRIOR APPLICATION NUMBER: US 60/285,408
: PRIOR FILING DATE: 2001-04-20
: PRIOR APPLICATION NUMBER: US 60/286,482
: PRIOR FILING DATE: 2001-04-25
: PRIOR APPLICATION NUMBER: US 60/341,050
: PRIOR FILING DATE: 2001-10-22
: PRIOR APPLICATION NUMBER: US 60/341,105
: PRIOR FILING DATE: 2001-10-22
: PRIOR APPLICATION NUMBER: US 09/895,834
: PRIOR FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: US 60/285,424
: PRIOR FILING DATE: 2001-04-20
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: FastSeq for Windows version 4.0
: SEQ ID NO 11
: LENGTH: 202

```

[illegible]

```

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8136
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Streptomyces avermectilis
; US-10-156-761-8136

```

```

Query Match
Best Local Similarity 5.1%; Score 69; DB 15; Length 149;
Matches 25; Conservative 12; Mismatches 24; Indels 20; Gaps 3;

```

```

QY 168 LITGELTTL-----ASRQQLIDMMEADKVAAPLRSALPAGWFIADKSG----- 211
DB 50 LLDLITLTIKRLILVASYDKAKKEMGIDMWEHD---PALSSHADGGRELAENHRLREEI 105
QY 212 AGERGSGIITAAALGPDGKPSR 232
DB 106 AGLRDAQALPSAEGPAEKPER 126

```

```

RESULT 9
US-10-233-926-4
; Sequence 4, Application US/10/233926
; Publication No. US20030131382A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kihney, Anthony J.
; TITLE OF INVENTION: PLANT CHOLINE PHOSPHATE CYTIDYLTRANSFERASE
; FILE REFERENCE: BB1419 US NA
; CURRENT APPLICATION NUMBER: US/10/233,926
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: US/09/735,846
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (18)
; US-10-233-926-4

```

```

Query Match
Best Local Similarity 5.0%; Score 67.5; DB 16; Length 149;
Matches 22; Conservative 23; Mismatches 33; Indels 25; Gaps 4;

```

```

QY 159 VAMATTLRLKLL-----TGELTTLASRQQLIDMMEADKVAAPLRSALPAGWFIADKSGA 212
DB 14 LSLTSLPLNLAMADHAAEAAPSSQEEEDMKAEAGDGDV-----EYADRGGG 64
QY 213 GERSGIIITAAALGPDGKPSRIVV-----IYTTGSOATMDERNR 250
DB 65 GGAANGCI-----PEGRRIRYADGIYDLFHFHGAHKSLEQARR 102

```

```

RESULT 10
US-09-948-018-8
; Sequence 8, Application US/09948018
; Patent No. US20020150977A1
; GENERAL INFORMATION:
; APPLICANT: Theill et al
; TITLE OF INVENTION: TNF RECEPTOR-LIKE MOLECULES AND USES THEREOF

```

```

; FILE REFERENCE: 01017/37677
; CURRENT APPLICATION NUMBER: US/09/948,018
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/230,191
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-948-018-8

```

```

Query Match
Best Local Similarity 5.0%; Score 67.5; DB 10; Length 187;
Matches 41; Conservative 22; Mismatches 57; Indels 63; Gaps 9;

```

```

QY 112 LITTTGGPKELTAPLHNMGDHTRLDRNEPE-----LNEAIPNDERDITTPVAMAT- 163
DB 11 LLLALAPPPAS-----QYCGRLTYWNPNNKCCSCSLQRRGP-----PPCGALETG 58
QY 164 -----TLRKLITGELTFLASR--QQLLDMMA-----DKVAGP-----L 195
DB 59 DTKKEASLPLSLRSLSSASQPLSLRLDELVLELIVLDPEPGGGGMAGTTRHLLA 118
QY 196 LRSALPAGWFIADKSGAGERSGIIITAAALGPDGKPSR--IYVYTTGSOATMDERNROIA 253
DB 119 ARYGLPAAW-----STRAYSILRPSRSLRLALILEWVAVAREPSASLQOLGTHLA 165
QY 254 EIG 256
DB 166 QLG 168

```

```

RESULT 11
US-09-738-626-5680
; Sequence 5680, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKIO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5680
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-5680

```

```

Query Match
Best Local Similarity 5.0%; Score 67; DB 10; Length 178;
Matches 25; Conservative 17; Mismatches 37; Indels 22; Gaps 4;
QY 171 GELTTLASRQQLIDW-----MEADKVAAPLRSALPAGWFIADKSGAGERSGIIIA 223

```

```

US-10-144-259-8
? Sequence 8, Application US/101444259
? Publication No. US20030109691A1
? GENERAL INFORMATION:
? APPLICANT: Arnaout, M. Amin
? APPLICANT: Li, Rui
? APPLICANT: Xiong, Jian-Ping
? TITLE OF INVENTION: VARIANT INTEGRIN POLYPEPTIDES AND USES THEREOF
? FILE REFERENCE: 00786-548001
? CURRENT APPLICATION NUMBER: US/10/144,259
? CURRENT FILING DATE: 2002-09-04
? PRIOR APPLICATION NUMBER: US 09/758,493
? PRIOR FILING DATE: 2001-01-11
? PRIOR APPLICATION NUMBER: US 60/221,950
? PRIOR FILING DATE: 2000-07-31
? NUMBER OF SEQ ID NOS: 30
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 8
? LENGTH: 193
? TYPE: prt
? ORGANISM: Homo sapiens
US-10-144-259-8

```

| | | | | |
|-----------------------|-------|---|----------------|-------------|
| Query Match | 5.2% | Score 70.5: | DB 15, | Length 193; |
| Best Local Similarity | 24.2% | Pred. NO. 34: | | |
| Matches | 31; | Conservative 15; | Mismatches 35; | Indels 47; |
| | | | Gaps 5 | |
| OY | 59 | IDAGQELGRHRYHSQNDLVESP--VTEKHLDGATVRELCSAATMSDNTAANLLLT | 116 | |
| | | | | |
| Db | 36 | IGRGQGVG-----VVGQGVDFVIEFHNDYRSKQVVEAASHIEQR----- | 77 | |
| OY | 117 | IGGPKELTALPHNNGDHYRLDRWE-----PELNEALPNDERD | 154 | |
| | | | | |
| Db | 78 | -GGTEHTTAP---GIEFARSEAFQKGRKAKVMIVITDGESHSDPLEKVIQOOSRD | 132 | |
| OY | 155 | TTTPVAMA 162 | | |
| | | | | |
| Db | 133 | NTRYAYVA 140 | | |

```

      RESULT 5
US-10-156-761-13867
; Sequence 13867, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13867
; LENGTH: 206
; TYPE: PRN
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13867

```

```

Query Match      5.28; Score 70; DB 15; Length 206;
Best local Similarity 25.28; Pred. No. 42;
Matches 53; Conservative 24; Mismatches 73; Indels 60; Gaps 11

QY    70 IHYSQNDLVESPYTEKHITDGMTV-----RELCSAATMSMDNTANLLLTIT---G 118
      :| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db      1 VHASQG---AOGPGTESAGEEKVTMDAAGESEFEVA-----NRSSALKTAVLLSG 50
QY      119 GRE-----LTAFLHMGDHWIRLD-----RW-----EPLEINAI PNDE 154
Db      51 GDRHAEDLLONALIKADWRSRIDPEEAVYRGVLYPROOVSRWLRKWRRLRSVAP-PE 109
QY      153 RDRTTVPAAATTLRLKLTGELLTIASROOLI-----DMEADEKVGAPLRSALPAGWF 205
Db      110 ASTGPPDASAAELRLVYMGALARTLTAKORVLYLRYEDLPDAD--VARIICGSGTVRS 167
QY      206 IADKSGAGERSGIIAALGP---DGKPSR 232
Db      168 TTHRSLARLKLAPELALGPDADAEQPSR 197

```

```

RESULT 7
US-10-101-464A-765
; Sequence 765, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/728,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 765
; LENGTH: 184
; TYPE: prt
; ORGANISM: Pinus radiata
US-10-101-464A-765

```

| Query Match | Similarity | 5.2% | Score 69.5 | DB 15 | Length 184 |
|-------------|--------------|--|--------------|-------|-------------------|
| Best local | Similarity | 24.7% | Pred. No. 40 | | |
| Matches | Conservative | 22 | Mismatches | 64 | Indels 45; Gaps 9 |
| QY | 41 | EPNMSPEKVL-----LCGAVLSRIDAGGOLGRIRHYSQNDLYEYSPV-----TEKHILNDG | 91 | | |
| | | : : : : : : : : : : : : : : | | | |
| Db | 23 | FCMTSTERLLVYPYMLNGSVASRL-----RDSINGKRALDMPTRKRISLG | 67 | | |
| QY | 92 | -----MTVREICSAITMSDNTAANLL-----TTGGPKELTAFLHMGDHVTRLDRLM-- | 140 | | |
| | | : : : : : : : : : : : : : : : : : | | | |
| Db | 68 | NANGLLIHEQCDDCKIIRHDYKRAANILLDEYFEAVVGDGFIAKLLDRDSHVTTAAVRCSTV | 127 | | |
| QY | 141 | ---EPLEINELPNDERDITTPYAMMTTKLLTIGEL---LTLASROO--LIDMM | 186 | | |
| | | : : : : : : : : : : : : : : : | | | |
| Db | 128 | GHAIPELYLSTGSSSK--TIDVFGFGLLLELITLGGALDFGRANOGVAMLDNV | 179 | | |

RESULT 8
 US-10-156-761-8136
 ; Sequence 8136, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SUTBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA

QY 135 TRLDWEPELNEAIPNDERDTTPVAMATT 164
|||
Db 61 TRLDWEPELNEAIPNDERDTTPVAMATT 90

RESULT 2

US-10-127-816-9
; Sequence 9, Application US/10127816
; Publication No. US20030104416A1

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Fox, Brian A.

APPLICANT: Klucher, Kevin M.

APPLICANT: Taft, David W.

APPLICANT: Kindsvogel, Wayne R.

TITLE OF INVENTION: CYTOKINE PROTEIN FAMILY

FILE REFERENCE: 01-17

CURRENT APPLICATION NUMBER: US/10/127,816

PRIOR FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: US 60/285,408

PRIOR FILING DATE: 2001-04-20

PRIOR APPLICATION NUMBER: US 60/286,482

PRIOR FILING DATE: 2001-04-25

PRIOR APPLICATION NUMBER: US 60/341,050

PRIOR FILING DATE: 2001-10-22

PRIOR APPLICATION NUMBER: US 60/341,105

PRIOR FILING DATE: 2001-10-22

PRIOR APPLICATION NUMBER: US 09/895,834

PRIOR FILING DATE: 2001-06-29

PRIOR APPLICATION NUMBER: US 60/285,424

PRIOR FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 59

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9

LENGTH: 202

TYPE: PRT

ORGANISM: Homo sapiens

US-10-127-816-9

Query Match 5.6%; Score 75; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 12;

Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

QY 50 LLCGAVSRIDAGEQLGRRIHYSND--LVESPYTEKHL-----TGQMTYREL----- 97

Db 17 LLAAVLTTRQADVPRATRLPVEAKDCHIAQFKSLSPRELOAFKRAKDAIERKLEKDL 76

QY 98 -CSA-----ATMS--DNTAANLLTTIGSPKELTAFLAN 129

Db 77 RCSSHLFPRAWMDKLQVQERPKALQAEVALTLKWNMTDSALATILGQPLHTLSHIS 136

QY 130 MGDHVT-----RLDRWEPELNEAIPNDERDTTPVAMATTLLKLLTGELLTL 176

Db 137 QLQTCYQLQATAEPSPSRRLSRMLRLQEA-QSKETPCGLASVTSNLFRLTLRLKCV 195

QY 177 ASRQQLI 183

Db 196 ANGDQCV 202

RESULT 3

US-10-142-717-12

; Sequence 12, Application US/10142717

; Publication No. US20030104579A1

GENERAL INFORMATION:

APPLICANT: Immunex Corporation

APPLICANT: Baum, Peter R.

APPLICANT: Mosley, Bruce A.

APPLICANT: Ketchum, Randal R.

APPLICANT: Taylor, Scott L.

TITLE OF INVENTION: CYTOKINE POLYPEPTIDES

FILE REFERENCE: 3282-A

CURRENT APPLICATION NUMBER: US/10/142,717

CURRENT FILING DATE: 2002-05-08

NUMBER OF SEQ ID NOS: 18

SOFTWARE: PatentIn version 3.1

SEQ ID NO 12

LENGTH: 202

TYPE: PRT

ORGANISM: Mus musculus

US-10-142-717-12

Query Match 5.6%; Score 75; DB 15; Length 202;

Best Local Similarity 21.4%; Pred. No. 12;

Matches 40; Conservative 26; Mismatches 67; Indels 54; Gaps 7;

QY 50 LLCGAVSRIDAGEQLGRRIHYSND--LVESPYTEKHL-----TGQMTYREL----- 97

Db 17 LLAAVLTTRQADVPRATRLPVEAKDCHIAQFKSLSPRELOAFKRAKDAIERKLEKDL 76

QY 98 -CSA-----ATMS--DNTAANLLTTIGSPKELTAFLAN 129

Db 77 RCSSHLFPRAWMDKLQVQERPKALQAEVALTLKWNMTDSALATILGQPLHTLSHIS 136

QY 130 MGDHVT-----RLDRWEPELNEAIPNDERDTTPVAMATTLLKLLTGELLTL 176

Db 137 QLQTCYQLQATAEPSPSRRLSRMLRLQEA-QSKETPCGLASVTSNLFRLTLRLKCV 195

QY 177 ASRQQLI 183

Db 196 ANGDQCV 202

RESULT 4

US-09-805-354-8

; Sequence 8, Application US/09805354

; Publication No. US20030078375A1

GENERAL INFORMATION:

APPLICANT: Atmaout, M. Amin

APPLICANT: Li, Rui

APPLICANT: Xiong, Jian-Ping

TITLE OF INVENTION: HIGH AFFINITY INTEGRIN POLYPEPTIDES AND USES THEREOF

FILE REFERENCE: 00786-536001

CURRENT APPLICATION NUMBER: US/09/805,354

CURRENT FILING DATE: 2002-06-04

PRIOR APPLICATION NUMBER: US 09/758,493

PRIOR FILING DATE: 2001-01-11

PRIOR APPLICATION NUMBER: US 60/221,950

PRIOR FILING DATE: 2000-07-31

NUMBER OF SEQ ID NOS: 20

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 8

LENGTH: 193

TYPE: PRT

ORGANISM: Homo sapiens

US-09-805-354-8

Query Match 5.2%; Score 70.5; DB 11; Length 193;

Best Local Similarity 24.2%; Pred. No. 34;

Matches 31; Conservative 15; Mismatches 35; Indels 47; Gaps 5;

QY 59 IDAGEQLGRRIHYSNDLVESPYTEKHL-----TGQMTYREL-----TGQMTYREL----- 116

Db 36 IGPGLQVQ-----VVOGEDVHFEHLNRYRSKDVAAASHLEQR----- 77

QY 117 IGPKEKLTAFLLAMGDHVTRLDRWE-----DELNEAIPNDERD 154

Db 78 -GGTEFRTAF-----GIEFARSEAFKGRGAKVMIVITDGESSHSDPLEKVIQOOSRD 132

QY 155 TTTPVAMA 162

Db 133 NNTRYAVA 140

RESULT 5

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 10, 2003, 12:20:56 ; Search time 17.5714 Seconds
(without alignments)
2183.941 Million cell updates/sec

Title: SRO2_157T
Perfect score: 1348
Sequence: 1 HPETLVKVKAEADQLGARVG.....TMDERNKQIAEIGASLTKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 541936 seqs, 145912426 residues

Total number of hits satisfying chosen parameters: 278642

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 457 | 33.9 | 94 | US-10-102-806-760 | Sequence 760, App |
| 2 | 75 | 5.6 | 202 | US-10-127-816-9 | Sequence 9, Appl1 |
| 3 | 75 | 5.6 | 202 | US-10-142-717-12 | Sequence 12, Appl1 |
| 4 | 70.5 | 5.2 | 193 | US-09-805-354-8 | Sequence 8, Appl1 |
| 5 | 70.5 | 5.2 | 193 | US-10-144-259-8 | Sequence 8, Appl1 |
| 6 | 70 | 5.2 | 206 | US-10-156-761-13867 | Sequence 13867, A |
| 7 | 69.5 | 5.2 | 184 | US-10-101-464A-765 | Sequence 765, App |
| 8 | 69 | 5.1 | 149 | US-10-233-926-4 | Sequence 8136, App |
| 9 | 67.5 | 5.0 | 149 | US-10-156-761-8136 | Sequence 8136, App |
| 10 | 67.5 | 5.0 | 187 | US-09-948-018-8 | Sequence 8, Appl1 |
| 11 | 67 | 5.0 | 178 | US-09-738-626-5680 | Sequence 5680, App |
| 12 | 67 | 5.0 | 194 | US-10-156-761-8623 | Sequence 8623, App |
| 13 | 66 | 4.9 | 202 | US-10-127-816-11 | Sequence 11, Appl1 |
| 14 | 64.5 | 4.8 | 191 | US-10-156-761-8978 | Sequence 8978, App |
| 15 | 64.5 | 4.8 | 195 | US-10-156-761-12656 | Sequence 12656, A |

| | | | | | |
|----|------|-----|-----|---------------------|--------------------|
| 16 | 64 | 4.7 | 162 | US-09-738-626-4796 | Sequence 4796, App |
| 17 | 64 | 4.7 | 179 | US-09-764-868-757 | Sequence 757, App |
| 18 | 64 | 4.7 | 179 | US-10-106-698-4858 | Sequence 4858, App |
| 19 | 64 | 4.7 | 202 | US-10-189-346-12 | Sequence 12, Appl1 |
| 20 | 63.5 | 4.7 | 83 | US-10-100-252-6 | Sequence 6, Appl1 |
| 21 | 63.5 | 4.7 | 144 | US-10-131-406-4 | Sequence 4, Appl1 |
| 22 | 63.5 | 4.7 | 189 | US-09-815-242-5862 | Sequence 5862, A |
| 23 | 63.5 | 4.7 | 189 | US-09-815-242-12979 | Sequence 12979, A |
| 24 | 63.5 | 4.7 | 206 | US-09-738-626-5425 | Sequence 5425, App |
| 25 | 62.5 | 4.6 | 146 | US-10-101-464A-516 | Sequence 516, App |
| 26 | 62.5 | 4.6 | 190 | US-10-156-761-9507 | Sequence 9507, App |
| 27 | 62.5 | 4.6 | 191 | US-09-764-870-390 | Sequence 390, App |
| 28 | 62.5 | 4.6 | 191 | US-10-125-540-390 | Sequence 390, App |
| 29 | 62 | 4.6 | 174 | US-09-864-761-35777 | Sequence 35777, A |
| 30 | 62 | 4.6 | 191 | US-09-927-827-64 | Sequence 64, Appl1 |
| 31 | 62 | 4.6 | 202 | US-10-156-761-13460 | Sequence 13460, A |
| 32 | 61.5 | 4.6 | 142 | US-09-862-027-31 | Sequence 31, Appl1 |
| 33 | 61.5 | 4.6 | 196 | US-09-738-626-6624 | Sequence 6624, App |
| 34 | 61.5 | 4.6 | 200 | US-10-156-761-12086 | Sequence 12086, A |
| 35 | 61 | 4.5 | 189 | US-09-764-870-523 | Sequence 523, App |
| 36 | 61 | 4.5 | 189 | US-10-125-540-523 | Sequence 523, App |
| 37 | 60.5 | 4.5 | 160 | US-09-882-227-230 | Sequence 230, App |
| 38 | 60.5 | 4.5 | 162 | US-10-141-531-5 | Sequence 5, Appl1 |
| 39 | 60.5 | 4.5 | 202 | US-10-189-346-16 | Sequence 16, Appl1 |
| 40 | 60 | 4.5 | 190 | US-09-738-626-4637 | Sequence 4637, App |
| 41 | 59.5 | 4.4 | 182 | US-10-156-761-12056 | Sequence 12056, A |
| 42 | 59.5 | 4.4 | 196 | US-10-219-220-265 | Sequence 265, App |
| 43 | 59.5 | 4.4 | 206 | US-09-741-669-461 | Sequence 461, App |
| 44 | 59.5 | 4.4 | 206 | US-09-815-242-10335 | Sequence 10335, A |
| 45 | 59 | 4.4 | 111 | US-09-867-550-1868 | Sequence 1868, App |

ALIGNMENTS

RESULT 1
US-10-102-806-760
Sequence 760, Application US/10102806
Publication No. US20030054421A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA103PIC1
CURRENT APPLICATION NUMBER: US/10/102,806
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: 09/925,298
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05881
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 846
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 760
LENGTH: 94
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (80)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (91)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-102-806-760

Query Match 33.9%: Score 457; DB 15; Length 94;
Best Local Similarity 97.8%: Pred. No. 8.5e-40;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 75 NDIVESPTTEKILTDGMTVRELCSAAITNSDNTAANLLTTIGGPKELTAPLHNGDHV 134
DB 1 NDIVESPTTEKILTDGMTVRELCSAAITNSDNTAANLLTTIGGPKELTAPLHNGDHV 60

;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 22681
;; LENGTH: 132
;; TYPE: PRF
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22681

Query Match
Best Local Similarity 29.8%; Score 62; DB 4; Length 132;
Matches 28; Conservative 8; Mismatches 34; Indels 24; Gaps 4;

QY 144 LNEAIPNDERDTTVPAMATTLRLKLTGELLTLASRQGLIDWNEADKVPGLRSALP-- 201
DB 34 LRGQIPN--HAGRRQQAATVLPMEPNE-----RRVAPMLKSPPPR 74
QY 202 AGWFIADKSGAGERSRGIIA---ALGPDGKPSR 232
DB 75 AAPEVADRAGPGPGTGTGCCAPRRAPAGVAPAR 108

RESULT 14
US-09-345-473E-31
; Sequence 31, Application US/09345473E
; Patent No. 6558903
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin
; TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof
; FILE REFERENCE: 35800/183781
; CURRENT APPLICATION NUMBER: US/09/345,473E
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 142
; TYPE: PRF
; ORGANISM: Rattus norvegicus
US-09-345-473E-31

Query Match
Best Local Similarity 41.7%; Score 61.5; DB 4; Length 142;
Matches 15; Conservative 5; Mismatches 11; Indels 5; Gaps 1;

QY 210 SGAGERSRGIIALGP-----DQPSRIIVYITG 240
DB 47 SGGSRDSGGTLYAIPELDNDGKASKADVYSFG 82

RESULT 15
US-09-252-991A-24737
; Sequence 24737, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24737
; LENGTH: 160
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24737

Query Match
4.6%; Score 61.5; DB 4; Length 160;

Best Local Similarity 37.5%; Pred. No. 70;
Matches 18; Conservative 6; Mismatches 19; Indels 5; Gaps 1;

QY 3 ETLVYKVDARDOLGARVGYIELDINSKRI-----ESFRPERPPMMS 45
DB 109 EILEKVKSLKSGAALVHLEVDENNPRAAVSFYKKSGEFFRERFYMMS 156

Search completed: September 10, 2003, 12:31:03
Job time: 11.8571 secs

Query Match 4.7%; Score 63.5; DB 2; Length 144;
Best Local Similarity 29.5%; Pred. No. 36;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

QY 66 LGRRIHY--SOND--LVEY--SP--VTEKHLTDGMTVRELCSAATMSDNTAANLLTTIG 118
DB 14 LGRLIHYNOKKDRLLNEYSPLDITTAQFVLCIR--CAACT-----56

QY 119 GKPELTAFLEH--NMGDHYTRLDR-----WEPELNEAIPNDERD-----TTTPVAMATTLRK 167
DB 57 -PEELKKVLSVDGLALRMLDRLVCKGWVERLPN--PNDKRGVLYKLTGGAAICEGCHQ 113

QY 168 LTTGELLTLASROOLIMMEADKVA--GPLLSALP 201
DB 114 LVGQDL-----HOLTKNLTRADEVATLEYLLKKVLP 144

RESULT 11
US-09-118-445-4
; Sequence 4, Application US/09118445
; Patent No. 6391545
; GENERAL INFORMATION:
; APPLICANT: LEVY, STUART B.
; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
; ASSAYS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0; Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/118,445
; FILING DATE: 17-JUL-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/225,480
; FILING DATE: <unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: T0359/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-118-445-4

Query Match 4.7%; Score 63.5; DB 4; Length 144;
Best Local Similarity 29.5%; Pred. No. 36;
Matches 46; Conservative 16; Mismatches 49; Indels 45; Gaps 12;

QY 66 LGRRIHY--SOND--LVEY--SP--VTEKHLTDGMTVRELCSAATMSDNTAANLLTTIG 118
DB 14 LGRLIHYNOKKDRLLNEYSPLDITTAQFVLCIR--CAACT-----56

QY 119 GKPELTAFLEH--NMGDHYTRLDR-----WEPELNEAIPNDERD-----TTTPVAMATTLRK 167
DB 57 -PEELKKVLSVDGLALRMLDRLVCKGWVERLPN--PNDKRGVLYKLTGGAAICEGCHQ 113

QY 168 LTTGELLTLASROOLIMMEADKVA--GPLLSALP 201
DB 114 LVGQDL-----HOLTKNLTRADEVATLEYLLKKVLP 144

RESULT 12
US-09-413-814-87
; Sequence 87, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 87
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-87

Query Match 4.6%; Score 62.5; DB 3; Length 198;
Best Local Similarity 22.4%; Pred. No. 76;
Matches 62; Conservative 29; Mismatches 79; Indels 107; Gaps 15;

QY 1 HPELVVYKDAEDDLGARVGYIELDNSKILIESPREFRPMSTF-----KVLGAV 55
DB 7 NPEAVYDAVKDTS--AAEVG---DATVHKVLEGIGAOVETAIWFGHEPESVLVAS 60

QY 56 LSRIDAGOEQGRRIHYSONDLVEYSPVTEKHLTDGMTVRELCSAATMSDNTAANLLTT 115
DB 61 LVR-----MGRIIARAA--TDROAD--ILR 82

QY 116 TIGGPKELTAFLEHNNGDHYTRLDRMEPELNEAIPNDERDTPPVAMATTLRKLTGELLT 175
DB 83 AVG-----ATRIYQLETEMGRVY--GADITWPLAQ-----DLDL 114

QY 176 LASROOLIMMEADKVAGPLLSALPACWFIADK--SCAGER-----GSRGIIAALG 225
DB 115 LASHYRVVPM-----NAGPLVGQTL--AGSKIRQRYRINIVGVPHTNRKRGDCKRLAPPT 169

QY 226 PDGKPSRIIVYITTSQATM-----DERNRQIAETG 256
DB 170 PD-----YVIRDGDTLLVSDSDVSRRVAEVG 197

RESULT 13
US-09-252-991A-22681
; Sequence 22681, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788

```

1 / REGISTRATION NUMBER: 35,024
2 / REFERENCE/DOCKET NUMBER: 0228.95
3 / TELECOMMUNICATION INFORMATION:
4 / TELEPHONE: 301-504-6629
5 / TELEFAX: 301-504-5060
6 / INFORMATION FOR SEQ ID NO: 2:
7 / SEQUENCE CHARACTERISTICS:
8 / LENGTH: 203 amino acids
9 / TYPE: amino acid
10 / TOPOLOGY: linear
11 / MOLECULE TYPE: protein
12 / US-08-624-677A-2

Query Match 4.7%; Score 64; DB 4; Length 203;
Best Local Similarity 23.2%; Pred. No. 53;
Matches 38; Conservative 22; Mismatches 74; Indels 30; Gaps 6;

QY 54 AVLSRIDAGEGEGRIHYSNDLVESPYVEKHLTDGTVBELCSAA--ITWSDNTAAN 111
DB 33 ACVSNTYDGDGDDAAGNPDV-----SDVTDDATTDDEMPHVSQCKPHTTGKSLIRK 83
QY 112 LLTTTIGGPELTAFLINMGDHV-TRLDKWEPELNEAIPNDRDPT-----T 137
DB 84 LAVPVYGA--LTSYL--VADRVLPELTSABEGCTESIPGKKRVATGIALVAAAAFA 138
QY 158 PVAATTLKRLTGLTGLTLLTASRQQLIDMEADKVAGPLRSALP 201
DB 139 GGLATFTFRHFYVPRKSKTVASSEDGALGNSEQYVGVNGSSDP 182

RESULT 10
US-08-225-480-4
; Sequence 4, Application US/08225480
; Patent No. 5817793
; GENERAL INFORMATION:
; APPLICANT: LEVY, STUART B.
; TITLE OF INVENTION: MULTIPLE ANTIBIOTIC RESISTANCE OPERON
; TITLE OF INVENTION: ASSAYS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,480
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938,085
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.
; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: T0359/7003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-225-480-4

```


Db 12 MGDGIMEREHGTGNTQCPFLRVVPGDESPILDORASEDILIT--NLSPITKSHPPS 69
Oy 181 OLIDWMEA-DKXAGPL-----RSALPAGFIADKSGAGRGSGITIAALGPDKPSR 232
Db 70 RDDWVEPDALOGTAVFATGDKATPFAVGI-----GASTRGS-GILASISPFOPAR 123

RESULT 2

US-09-252-991A-24241
Sequence 24241, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 24241
LENGTH: 197
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24241

Query Match
Best Local Similarity 5.4%; Score 72.5; DB 4; Length 197;
Matches 37; Conservative 23; Mismatches 42; Indels 55; Gaps 7;

Oy 94 VRELSAATITSDNTANLLTTIGGPKELFAFLNMGDVTPLDWEPELNEAIPDER 153
Db 12 VAKLSAATITSDNTANLLTTIGGPKELFAFLNMGDVTPLDWEPELNEAIPDER 153
Oy 154 DTTTPVA-MATTLKRLTGELTLA-----SRQOLIDWEADVAGP----- 194
Db 56 ATTSPLACRACRTRSLSCG--LARANSTPGSTSRRR--WSSSNSISAVSAGRSMPI 110
Oy 195 -----LLRSALPAGFIADKSGAGRG 216
Db 111 SMAPARAVSTWSPVITFTALPAMHSAATATASRG 147

RESULT 3

US-08-858-207A-519
Sequence 519, Application US/08858207A
Patent No. 6348328
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
TITLE OF INVENTION: Stodola, Robert
NUMBER OF SEQUENCES: NO. 6348328el Compounds
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 519:
SEQUENCE CHARACTERISTICS:
LENGTH: 170 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-519

Query Match
Best Local Similarity 5.3%; Score 71.5; DB 4; Length 170;
Matches 46; Conservative 34; Mismatches 70; Indels 63; Gaps 8;

Oy 15 LGARVGIYELDINSKILFSPREPRPMSTFVLLCGAVLSRIDAG--POLGRRIHY 72
Db 1 LRNIGIVLQDLSTRELISIPCTKKSVMGRF-----AAAFVADSEIQLPORVD 54
Oy 73 SQNDLVEYSPVTEKHLDGHTVRELSAATITMSDNTANLLTTIGGPRELFAFLNMGD 132
Db 55 -----SPVSRGSSFSFGQROLAFARTVASO-----PRILY----- 86
Oy 133 HTYRLDRWPELNEAIPDERDITTPVAMATTLKRLTGELTLASROOLIDWEADVKA 192
Db 87 -----LDEAFANIDSESLV--QASLAKMRGR--TTIAIARLSTIODNCT- 131
Oy 193 GPLRSALPAGFIADKSGAGRGSGITIAALG 225
Db 132 -----YVLDKGRILIESGTHEELIALG 152

RESULT 4

US-08-991-890-4
Sequence 4, Application US/08991890
Patent No. 6114307
GENERAL INFORMATION:
APPLICANT: Jaspers, Stephen R.
APPLICANT: Sprugel, Katherine H.
APPLICANT: Ren, Hong Ping
APPLICANT: Humes, Jacqueline M.
APPLICANT: Hoffman, Ross C.
APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 1201 Eastlake Avenue East
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,890
FILING DATE:
CLASSIFICATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using SW model

Run on: September 10, 2003, 12:19:51 : Search time 11.8571 Seconds
(without alignments)
938.485 Million cell updates/sec

Title: SEQ2_157T
Perfect score: 1348
Sequence: 1 HPETLVKVKDAEDQLGARVQ.....TMDENRRIAEIGSLIKHW 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 263404

Minimum DB seq length: 0
Maximum DB seq length: 207

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--|
| 1 | 74.5 | 5.5 | 124 | 4 | US-08-311-731A-202 Sequence 202, App |
| 2 | 72.5 | 5.4 | 197 | 4 | US-09-252-991A-24241 Sequence 24241, A |
| 3 | 71.5 | 5.3 | 170 | 4 | US-08-858-207A-519 Sequence 519, App |
| 4 | 71 | 5.3 | 159 | 3 | US-08-991-890-4 Sequence 4, Appl |
| 5 | 69 | 5.1 | 169 | 4 | US-09-328-352-4491 Sequence 4491, Ap |
| 6 | 65.5 | 4.9 | 158 | 3 | US-09-010-809-19 Sequence 19, Appl |
| 7 | 65 | 4.8 | 150 | 3 | US-09-239-809-2 Sequence 2, Appl |
| 8 | 64.5 | 4.8 | 178 | 4 | US-09-252-991A-29942 Sequence 29942, A |
| 9 | 64 | 4.7 | 203 | 4 | US-08-624-677A-2 Sequence 2, Appl |
| 10 | 63.5 | 4.7 | 144 | 2 | US-08-225-480-4 Sequence 4, Appl |
| 11 | 63.5 | 4.7 | 144 | 3 | US-09-413-814-87 Sequence 87, Appl |
| 12 | 62.5 | 4.6 | 198 | 4 | US-09-252-991A-22681 Sequence 22681, A |
| 13 | 62 | 4.6 | 132 | 4 | US-09-345-473E-31 Sequence 31, Appl |
| 14 | 61.5 | 4.6 | 142 | 4 | US-09-252-991A-24737 Sequence 24737, A |
| 15 | 61.5 | 4.6 | 160 | 4 | US-09-252-991A-24737 Sequence 24737, A |
| 16 | 61 | 4.5 | 181 | 2 | US-08-482-142-195 Sequence 195, App |
| 17 | 61 | 4.5 | 181 | 2 | US-08-478-572-195 Sequence 195, App |
| 18 | 61 | 4.5 | 181 | 3 | US-08-484-296-195 Sequence 195, App |
| 19 | 61 | 4.5 | 189 | 4 | US-08-671-548C-48 Sequence 48, Appl |
| 20 | 60.5 | 4.5 | 103 | 4 | US-09-732-210-1282 Sequence 1282, Ap |
| 21 | 60.5 | 4.5 | 167 | 2 | US-08-690-849-2 Sequence 2, Appl |
| 22 | 60.5 | 4.5 | 167 | 3 | US-09-004-053-2 Sequence 2, Appl |
| 23 | 60 | 4.5 | 144 | 4 | US-09-252-991A-25578 Sequence 25578, A |
| 24 | 59.5 | 4.4 | 146 | 4 | US-09-134-001C-5269 Sequence 5269, Ap |
| 25 | 59.5 | 4.4 | 178 | 3 | US-09-220-731-24 Sequence 24, Appl |
| 26 | 59.5 | 4.4 | 178 | 4 | US-09-242-999-24 Sequence 24, Appl |
| 27 | 59 | 4.4 | 49 | 2 | US-08-284-465-9 Sequence 9, Appl |

| | | | | | |
|----|------|-----|-----|---|--|
| 28 | 59 | 4.4 | 145 | 4 | US-09-134-001C-4982 Sequence 4982, Ap |
| 29 | 59 | 4.4 | 171 | 4 | US-09-107-532A-3979 Sequence 3979, Ap |
| 30 | 59 | 4.4 | 174 | 2 | US-08-557-122A-6 Sequence 6, Appl |
| 31 | 59 | 4.4 | 174 | 4 | US-09-262-666-6 Sequence 6, Appl |
| 32 | 59 | 4.4 | 180 | 4 | US-09-194-905-5 Sequence 5, Appl |
| 33 | 59 | 4.4 | 199 | 4 | US-09-252-991A-30363 Sequence 30363, A |
| 34 | 59 | 4.4 | 200 | 2 | US-08-557-122A-12 Sequence 12, Appl |
| 35 | 59 | 4.4 | 200 | 4 | US-09-262-666-12 Sequence 12, Appl |
| 36 | 58.5 | 4.3 | 167 | 4 | US-08-328-352-6378 Sequence 6378, Ap |
| 37 | 58.5 | 4.3 | 203 | 4 | US-09-252-991A-24921 Sequence 24921, A |
| 38 | 58 | 4.3 | 136 | 4 | US-09-252-991A-30622 Sequence 30622, A |
| 39 | 58 | 4.3 | 168 | 1 | US-08-451-947-6 Sequence 6, Appl |
| 40 | 58 | 4.3 | 168 | 2 | US-08-424-826A-6 Sequence 6, Appl |
| 41 | 58 | 4.3 | 168 | 3 | US-08-928-694-6 Sequence 6, Appl |
| 42 | 58 | 4.3 | 168 | 4 | US-08-450-842-6 Sequence 6, Appl |
| 43 | 58 | 4.3 | 168 | 4 | US-08-451-390-6 Sequence 6, Appl |
| 44 | 58 | 4.3 | 168 | 5 | PCT-US91-06950-6 Sequence 6, Appl |
| 45 | 58 | 4.3 | 183 | 4 | US-09-252-991A-31347 Sequence 31347, A |

ALIGNMENTS

RESULT 1
US-08-311-731A-202
Sequence 202, Application US/08311731A

Patent No. 6583266

GENERAL INFORMATION:

APPLICANT: SMITH, DOUGLAS

INVENTOR: MAO, JEN-I

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

TITLE OF INVENTION: RELATING TO MYCOBACTERIA TUBERCULOSIS AND LAPRAE FOR

NUMBER OF SEQUENCES: 411

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/311,731A

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/720-2441

TELEFAX: 617/720-3500

INFORMATION FOR SEQ ID NO: 202:

SEQUENCE CHARACTERISTICS:

LENGTH: 124 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Mycobacterium leprae

Query Match 5.5%; Score 74.5; DB 4; Length 124;
Best Local Similarity 28.6%; Pred. No. 1.7;
Matches 34; Conservative 19; Mismatches 43; Indels 23; Gaps 8;

QY 130 MGDHV--TRLDWEPELN--EAIIPNDERDT---TPVNAVATTLRLKLTGELLTLASRQ 180

THIS PAGE BLANK (USPTO)


```
RESULT 14
ID AAM55524 standard; Protein; 200 AA.
XX
AC AAM55524;
XX
DT 02-JUL-1998 (first entry)
XX
DE H. pylori ORF 29ep10720_24432762_c3_39 cellular protein.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
XX inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
OS Helicobacter pylori.
XX
PN WO9737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97WO-US05223.
XX
PR 06-DEC-1996; 9605-0761318.
PR 29-MAR-1996; 9605-0625811.
PR 02-APR-1996; 9605-0758731.
PR 25-OCT-1996; 9605-0736905.
PR 28-OCT-1996; 9605-0738859.
XX
PA (ASTR ) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR MPI: 1997-50312/46.
DR N-PSDB; AAV24933.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claims 14,93; Page 726-727; 1145pp; English.
XX
CC This sequence is a H. pylori cellular protein.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.
XX
SO Sequence 200 AA;
XX
Query Match 6.0%; Score 81; DB 18; Length 200;
Best Local Similarity 25.9%; Pred. No. 5.2;
Matches 51; Conservative 30; Mismatches 44; Indels 72; Gaps 13;
XX
QY 22 IELDINGKILDSF---RPERPMMSTFKVLLGAVISRIDAQEOIGRRIRHYSQNDIV 78
DB 11 VLLDSDQSKMEFTFAIRAEKERPFSLF-----NRSRGSPTLKQWAKRENIILI 61
QY 79 ----EYSPTEKHLDGKMTVRELCSAITSNDTANLL--TTIGPKKL-TAFIHNMG 131
```

```
DB 62 DTNGEYSKETOR-----AMLLS-----NIVLPFT---PSQDLTEVLANML 99
QY 132 DHVTRLDRWPEPELINEAIPNDRDTTTPVAMATTLKRLT-----GELLTTLASQQLIDWME 187
DB 100 ERIEQLQ-----ELNE-----NLKALIVINMPITPTLKEKRALTEFK 138
QY 188 ----ADKVGAPLLRSAL 200
DB 139 ENNPSDRIT--LLESSL 153
XX
RESULT 15
ID AAU45138 standard; Protein; 158 AA.
XX
AC AAU45138;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #6034.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR MPI: 2001-616774/71.
DR N-PSDB; AAS59525.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 6333; 1069pp; English.
XX
CC Sequences AAU93105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
```

DB 1 TLVKVDAEDQDQARVGYIELDLSNGKILSPFRPRPMMSTFVLLCGAVLSRIDAQ 60
 QY 64 EQLGRI-----HYSQNDLV 78
 61 EQLGRITILRMWTLSYHQSKSIL 85

RESULT 12

AA60626 standard; Protein: 119 AA.

AC AAP60626;

DT 25-MAR-2003 (updated)
 DT 17-JUN-1991 (first entry)

DE Beta-urogastrone - beta-lactamase fusion protein from pUG2301.

KW Beta-urogastrone gene; hormone; salivary gland; ulcers; wounds;
 fusion protein; beta-lactamase.

OS Synthetic.

PH Key Location/Qualifiers
 FT Protein 1..63

FT /label= beta-lactamase

FT Misc-difference 64..66

FT /label= adaptor

FT Protein 67..119

FT /label= beta-urogastrone

DE3523634-A.

PD 09-JAN-1986.

PF 02-JUL-1985; 85DE-3523634.

PR 02-JUL-1984; 84JP-0137691.

PA (EART) EARTH CHEM CO LTD.

PI Aoki S, Ohgai H, Horinaka A, Hiramatsu H, Koumoto S, Nishimura A;

PI Matsushiro S;

XX MPI: 1986-015031/03.

DR N-PSDB; AAN60630.

XX New gene for expression of beta-urogastrone - its derivs., plasmid(s)
 PT and transformed cells contg. it.

PS Disclosure; Page 55; 92pp; German.

XX The fusion protein is less easily degraded by proteases and so
 CC protects beta-urogastrone and beta-lactamase collects in the periplasm
 CC of E. coli. It is therefore easy to collect and purify the product.

CC Beta-urogastrone is the hormone of the salivary glands which suppresses
 CC stomach acid secretion and promotes cell growth, so is useful for
 CC treating ulcers and wounds. Previously the product was obtd. only
 CC in small amts. from human urine.

CC See also AAN60628, and 30-32.

CC (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 119 AA:

Query Match 15.1%; Score 203; DB 7; Length 119;
 Best Local Similarity 100.0%; Pred. No. 4.3e-13;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPETLVKVKDAEDQLGARVGYIELDLSNGKILSPFRPR 40
 DB 24 HPETLVKVKDAEDQLGARVGYIELDLSNGKILSPFRPR 63

RESULT 13

AA92783 standard; Protein: 51 AA.

AC AA92783;

DT 29-AUG-2000 (first entry)

DE Vtgs-beta-lactamase fusion protein (partial).

KW Vtgs; vitellogenin; secretory signal sequence; gene expression;
 oestrogen receptor binding protein; systemic circulation;

KW beta-lactamase.

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

PH Key Location/Qualifiers
 FT Peptide 1..21

FT /label= Vtgs

FT Cleavage-site 15..16

FT Protein 22..51

FT /note= "beta-lactamase mature protein"

MO200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

PA (UYSI-) UNIV SINGAPORE NAT.

PA (LAMT/) LAM T J.

PI Ding JL, Tan NS, Ho B;

DR MPI: 2000-365615/31.

DR N-PSDB; AAA28507.

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

PS Example 6; Fig 14A; 73pp; English.

XX A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated pBADvtglactkana was
 CC constructed. Vtgs and variants that comprise conservative

CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding

CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining

CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

XX Sequence 51 AA:

Query Match 10.8%; Score 145; DB 21; Length 51;
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PETLVKVKDAEDQLGARVGYIELDLSNGK 30
 DB 23 PETLVKVKDAEDQLGARVGYIELDLSNGK 51